

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2002, 14:30:48 ; Search time 34 Seconds  
(without alignments)  
231.229 Million cell updates/sec

Title: US-09-424-815E-1

Perfect score: 304  
Sequence: 1 KVHSLARAGKVRGQTPKVA.....RRFVVVPTFGKKKGNANS 59

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A.Geneseq\_101002.\*  
1: /SID2/gcgcdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SID2/gcgcdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SID2/gcgcdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SID2/gcgcdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SID2/gcgcdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SID2/gcgcdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SID2/gcgcdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SID2/gcgcdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SID2/gcgcdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SID2/gcgcdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SID2/gcgcdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SID2/gcgcdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SID2/gcgcdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SID2/gcgcdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SID2/gcgcdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SID2/gcgcdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SID2/gcgcdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SID2/gcgcdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SID2/gcgcdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SID2/gcgcdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SID2/gcgcdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SID2/gcgcdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SID2/gcgcdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	304	100.0	59	15	AA1980204
2	304	100.0	59	20	AA1985381
3	304	100.0	66	23	ABP42408
4	304	100.0	170	21	AA1943549
5	276	90.8	118	22	AB129449
6	276	90.8	118	22	AB134625
7	276	90.8	118	22	AB120035
8	276	90.8	118	22	AA155415
9	276	90.8	118	22	AA167811
10	276	90.8	118	22	AA15620

11	276	90.8	118	22	AA128118
12	276	90.8	118	22	AA103364
13	276	90.8	118	22	AB137354
14	263	86.5	132	22	AB170963
15	239	78.6	62	21	AA122753
16	239	78.6	62	21	AA132888
17	239	78.6	62	21	AA134119
18	239	78.6	62	21	AA14883
19	239	78.6	83	21	AA14882
20	239	78.6	84	21	AA132887
21	239	78.6	93	21	AA134118
22	239	78.6	110	21	AA132886
23	239	78.6	130	21	AA126728
24	238	78.3	62	21	AA140704
25	238	78.3	62	21	AA140788
26	238	78.3	82	21	AA140703
27	238	78.3	130	21	AA105569
28	231	76.0	62	21	AA104311
29	231	76.0	62	21	AA15348
30	231	76.0	62	21	AA138861
31	231	76.0	62	21	AA139476
32	231	76.0	62	21	AA148690
33	231	76.0	68	21	AA131871
34	211.5	69.6	312	22	AA130872
35	211	69.4	157	22	AA175965
36	208	68.4	41	17	AA103681
37	208	68.4	108	22	AB127228
38	208	68.4	108	22	AB127519
39	201	66.1	40	22	AA108731
40	188.5	62.0	58	21	AA181318
41	175	57.6	61	22	AB127230
42	175	57.6	90	22	AB127520
43	168.5	55.4	129	22	AA130219
44	163	53.6	35	22	AA107113
45	123.5	40.6	459	23	AB134132

#### ALIGNMENTS

RESULT 1  
AA1980204  
ID AA1980204 standard; protein; 59 AA.  
XX  
AC AA1980204;  
XX  
DT 12-JUL-1994 (first entry)  
XX  
DE Heparin-binding protein.  
XX  
KW Cell growth agent; wound; bone disease; treatment; agent.  
XX  
OS Rattus norvegicus.  
XX  
PN JP05339287-A.  
XX  
PD 21-DEC-1993.  
XX  
PF 05-JUN-1992; 92JP-0145125.  
XX  
PR 05-JUN-1992; 92JP-0145125.  
XX  
RA (FARH) HOECHST JAPAN LTD.  
XX  
DR WPI; 1994-031824/04.  
XX  
PT New heparin-binding protein - used as a cell growth agent for  
XX treatment of wounds and bone disease  
XX  
PS Claim 1; Page 3; 4pp; Japanese.  
XX  
CC The sequence is that of a heparin binding-protein which is useful as  
XX a cell growth agent and in the treatment of wounds and bone disease.

XX SQ Sequence 59 AA;  
Query Match 100.0%; Score 304; DB 15; Length 59;  
Best Local Similarity 100.0%; Pred. No. 2.5e-32;  
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KVHGLARAGKVRGOTPKVAKQEKKKKTGRAKRRMQYNRRFVNVVPTFGKKKGPNANS 59  
Db 1 KVHGLARAGKVRGOTPKVAKQEKKKKTGRAKRRMQYNRRFVNVVPTFGKKKGPNANS 59

RESULT 2  
AAW95381 ID AAW95381 standard; peptide; 59 AA.  
XX AC AAW95381;  
XX 17-MAR-1999 (first entry)  
XX Antimicrobial peptide fragment from ubiquicidine.  
XX Ubiquicidine; treatment; diagnosis; prophylaxis; infection; microbial;  
XX pathogenic; Gram-positive bacteria; antimicrobial; Staphylococcus aureus;  
XX Listeria monocytogenes; Gram-negative; Klebsiella pneumoniae; E. coli;  
XX enterococcus; Salmonella typhimurium; Mycobacterium avium; M. fortuitum;  
XX fungus; Candida albicans; Cryptococcus neoformans; Aspergillus fumigatus;  
XX virus; parasite; Trypanosoma cruzi; Taxoplasma gondii.  
XX Synthetic.  
XX Mus sp.  
XX WO9854314-A1.  
XX 03-DEC-1998.  
XX 29-MAY-1998; 98WO-NL00311.  
XX 29-MAY-1997; 97NL-1006164.  
XX (UYLE-) RIJKSUNIV LEIDEN.  
XX Feitsma RIJ, Hiemstra PS, Nibbering PH, Pauwels EKJ;  
XX Van Den Barselaar MT;  
XX WPI; 1999-070214/06.  
XX New antimicrobial peptides derived from ubiquicidine - useful for  
XX the prophylaxis, diagnosis and treatment of infections in humans and  
XX animals  
XX Claim 2; Page 23; 48pp; English.  
XX Sequences AAW95381-389 represent antimicrobial peptide fragments derived  
XX from ubiquicidine. Ubiquicidine or optionally modified peptide fragments  
XX of ubiquicidine, may be used for the treatment, diagnosis, or  
XX prophylaxis of infections in humans and animals. In particular the  
XX products and methods are directed against microbial infections caused by  
XX pathogenic Gram-positive Staphylococcus aureus, including antibiotic  
XX resistant strains, Listeria monocytogenes, and Gram-negative antibiotic  
XX resistant Klebsiella pneumoniae, E. coli, enterococci, and Salmonella  
XX typhimurium bacteria, micro-organisms difficult to treat such as  
XX Mycobacterium avium and M. fortuitum, fungi such as Candida albicans,  
XX Cryptococcus neoformans, and Aspergillus fumigatus, viruses, in  
XX particular enveloped viruses, and parasites such as Trypanosoma cruzi and  
XX Taxoplasma gondii.

XX SQ Sequence 59 AA;  
Query Match 100.0%; Score 304; DB 20; Length 59;  
Best Local Similarity 100.0%; Pred. No. 2.5e-32;  
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KVHGLARAGKVRGOTPKVAKQEKKKKTGRAKRRMQYNRRFVNVVPTFGKKKGPNANS 59  
Db 1 KVHGLARAGKVRGOTPKVAKQEKKKKTGRAKRRMQYNRRFVNVVPTFGKKKGPNANS 59

RESULT 3  
ABP42408 ID ABP42408 standard; Protein; 66 AA.  
XX AC ABP42408;  
XX 22-AUG-2002 (first entry)  
XX Human ovarian antigen HOCWY79, SEQ ID NO:3540.  
XX Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;  
XX ovarian cancer; breast cancer; tumour; reproductive system disorder;  
XX infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;  
XX PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;  
XX inflammatory condition; immune disorder; blood disorder;  
XX cardiovascular disorder; respiratory disorder; neurological disorder;  
XX gastrointestinal disorder; urinary system disorder; drug screening;  
XX gene therapy; chromosome mapping; forensic analysis;  
XX antibody preparation; cytostatic; immunomodulatory; neuroprotective;  
XX antiinflammatory; gynaecological; reproductive.  
XX Homo sapiens.  
XX WO200200677-A1.  
XX 03-JAN-2002.  
XX 07-JUN-2001; 2001WO-US18569.  
XX 07-JUN-2000; 2000US-209467P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Birse CE, Rosen CA;  
XX WPI; 2002-147878/19.  
XX N-PSDB; ABQ55485.  
XX Isolated nucleic acid molecules encoding novel ovarian polypeptides,  
XX useful in the prevention, treatment and diagnosis of cancer (e.g.  
XX ovarian cancer), immune disorders, cardiovascular disorders and  
XX neurological diseases -  
XX Claim 11; SEQ ID No 3540; 2922pp; English.  
XX The invention relates to 2175 novel human ovarian antigens (ABP41054-  
XX ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also  
XX encompasses polypeptides 90% identical and polynucleotides 95% identical  
XX to the sequences of the invention. The invention additionally relates to  
XX recombinant vectors and host cells comprising human ovarian antigen  
XX polynucleotides, antibodies against human ovarian antigens, and the use  
XX of ovarian antigen polynucleotides and polypeptides in diagnosing,  
XX treating, prognosing or preventing various ovary and/or breast-related  
XX disorders. Such conditions include ovarian cancer and breast cancer, and  
XX metastatic tumours of ovarian or breast origin, reproductive system  
XX disorders (e.g., infertility, disorders of pregnancy, anovulation,  
XX polycystic ovary syndrome, ovarian cysts, and dysmenorrhea), endocrine  
XX disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic  
XX shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and  
XX vaginitis), immune disorders (e.g., congenital and acquired  
XX immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),  
XX blood-related disorders (e.g., anaemia), cardiovascular disorders,  
XX respiratory disorders, neurological disorders, gastrointestinal disorders  
XX and urinary system disorders. Ovarian antigen polypeptides and  
XX polynucleotides may also be used in screening for compounds which  
XX modulate ovarian antigen expression or activity. The polynucleotides may  
XX further be used for gene therapy, chromosome mapping, in the  
XX identification of individuals and in forensic analysis, and the

CC polypeptides may be used as food additives or to prepare antibodies  
CC useful in disease diagnosis, drug targeting and phenotyping. The present  
CC sequence represents a human ovarian antigen of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).

XX Sequence 66 AA;

Query Match 100.0%; Score 304; DB 23; Length 66;  
Best Local Similarity 100.0%; Pred. No. 2.8e-32;  
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKHGSLARAGKVRGQTPKVAKOEKKKTGRAKRRMQYRRFVNVVPTFGKKKGPYANS 59  
DB 8 KKHGSLARAGKVRGQTPKVAKOEKKKTGRAKRRMQYRRFVNVVPTFGKKKGPYANS 66

RESULT 4  
AAB43549

ID AAB43549 standard; Protein; 170 AA.

AC AAB43549;

DT 08-FEB-2001 (first entry)

DE Human cancer associated protein sequence SEQ ID NO:994.

XX Human; cancer associated gene; cancer antigen; detection; cancer;  
XX diagnosis; cytostatic; proliferative; vunerary; immunomodulator;  
XX antidiabetic; antiaschemic; antithrombotic; antiatheritic; antiviral;  
XX antiinflammatory; antihypertoid; antiallergic; antibacterial; cardiac;  
XX dermatologic; neuroprotective; thrombolytic; coagulant; nootropic;  
XX vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;  
XX immune disorder; haematopoietic cell disorder; autoimmune disorder;  
XX allergic reaction; graft versus host disease; organ rejection;  
XX haemostatic; thrombolytic; cardiovascular disorder; infection;  
XX neurological disease; drug screening.

OS Homo sapiens.

PN WO20055350-A1.

PD 21-SEP-2000.

PR 08-MAR-2000; 2000WO-US05882.

PR 12-MAR-1999; 99US-0124270.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM;

XX WPI; 2000-587533/55.

DR N-PSDB; AAC77758.

PT Novel isolated nucleic acids comprising sequences encoding peptides  
XX useful for treating or diagnosing e.g. cancer -

PS Claim 11; Page 1568-1569; 2352pp; English.

XX AAC77607 to AAC78448 encode the human cancer associated proteins given  
CC in AAB43398 to AAB44239. The proteins can have activities based on the  
CC tissues and cells the genes are expressed in. Example of activities  
CC include: cytostatic; proliferative; vunerary; immunomodulator;  
CC antidiabetic; antiaschemic; antithrombotic; antiatheritic;  
CC antiinflammatory; antihypertoid; antiallergic; antibacterial; antiviral;  
CC dermatologic; neuroprotective; cardiac; thrombolytic; coagulant;  
CC nootropic; vasotropic; antipsoriatic and antiangiogenic. The  
CC polynucleotides and polypeptides can be used for preventing, treating or  
CC ameliorating medical conditions and diagnosing pathological conditions.  
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from  
CC the present invention may be used to treat immune disorders by activating

CC or inhibiting the proliferation, differentiation or mobilisation of  
CC immune cells, to treat disorders of haematopoietic cells, autoimmune  
CC disorders, allergic reactions, graft versus host disease and organ  
CC rejection, modulate haemostatic or thrombolytic activity, modulate  
CC inflammation, cancers, cardiovascular disorders, neurological disease and  
CC bacterial or viral infections. The peptides, nucleotides, antibodies,  
CC agonists and antagonists may be also be used in drug screens. AAC78449 to  
CC AAC78457 and AAB44240 represent sequences used in the exemplification of  
CC the present invention.

XX Sequence 170 AA;

Query Match 100.0%; Score 304; DB 21; Length 170;  
Best Local Similarity 100.0%; Pred. No. 8.1e-33;  
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKHGSLARAGKVRGQTPKVAKOEKKKTGRAKRRMQYRRFVNVVPTFGKKKGPYANS 59  
DB 112 KKHGSLARAGKVRGQTPKVAKOEKKKTGRAKRRMQYRRFVNVVPTFGKKKGPYANS 170

RESULT 5  
AAB29449

ID AAB29449 standard; Peptide; 118 AA.

AC AAB29449;

DT 01-FEB-2002 (first entry)

DE Peptide #2100 encoded by breast cell single exon nucleic acid probe.

XX Human; microarray; single exon probe; gene expression; breast;  
XX disease; cancer.

OS Homo sapiens.

PN WO200157271-A2.

PD 09-AUG-2001.

PR 30-JAN-2001; 2001WO-US00662.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 23-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-496933/54.

DR 2001-496933/54.

PT New spatially-addressable set of single exon nucleic acid probes,  
XX useful for measuring gene expression in sample derived from human  
XX breast, comprises number of single exon nucleic acid probes -

PS Claim 27; SEQ ID NO 12417; 327pp + sequence listing; English.

XX The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human breast and BT 474 cells. The method involves contacting  
CC the probes with a collection of detectably labelled nucleic acids  
CC derived from mRNA of human breast, and then measuring the label  
CC bound to each probe of the microarray. The probes are useful for  
CC verifying the expression of regions of genomic DNA predicted to  
CC encode proteins. They are useful for gene discovery, and for  
CC determining predisposition and/or prognosing breast disease. Gene  
CC expression analysis is useful for assessing the toxicity of chemical  
CC agents on cells. The microarray of this invention presents a far greater

CC diversity of probes for measuring gene expression, with far less bias  
CC than expressed sequence tag microarrays. The method is suitable for  
CC rapid production of functional information from genomic sequence. The  
CC present sequence is a peptide encoded by a single exon nucleic acid  
CC probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

SQ Sequence 118 AA;  
Query Match 90.8%; Score 276; DB 22; Length 118;  
Best Local Similarity 88.9%; Pred. No. 2.5e-28;  
Matches 56; Conservative 1; Mismatches 2; Indels 4; Gaps 1;

QY 1 KVHGLARAGKVRGQTPKVAQEKK----KKKTGRKRMQYNNRRFVNVVPTFGKKKGP 56  
DB 56 KVHGLARAGKVRGQTPKVAQEKKKKKKKTGQATRMQYNNRRFVNVVPTFGKKKGP 115  
QY 57 ANS 59  
DB 116 ANS 118

RESULT 6  
ABB34625  
ID ABB34625 standard; Peptide; 118 AA.

AC ABB34625;

DT 04-FEB-2002 (first entry)

DE Peptide #2131 encoded by human foetal liver single exon probe.

DE Human; foetal liver; gene expression; single exon nucleic acid probe.

OS Homo sapiens.

PN WO200157277-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00659.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

PS WPI; 2001-483447/52.

PT Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human fetal liver -

PS Claim 27; SEQ ID NO 27260; 639pp + sequence listing; English.

CC The invention relates to a single exon nucleic acid probe for

CC measuring human gene expression in a sample derived from human foetal

CC liver. The single exon nucleic acid probes may be used for predicting,

CC measuring and displaying gene expression in samples derived from human

CC fetal liver. The present sequence is a peptide encoded by a single exon

CC nucleic acid probe of the invention.

CC Note: The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 118 AA;

Query Match 90.8%; Score 276; DB 22; Length 118;

Best Local Similarity 88.9%; Pred. No. 2.5e-28;

Matches 56; Conservative 1; Mismatches 2; Indels 4; Gaps 1;

QY 1 KVHGLARAGKVRGQTPKVAQEKK----KKKTGRKRMQYNNRRFVNVVPTFGKKKGP 56  
DB 56 KVHGLARAGKVRGQTPKVAQEKKKKKKKTGQATRMQYNNRRFVNVVPTFGKKKGP 115  
QY 57 ANS 59  
DB 116 ANS 118

RESULT 7  
ABB20035  
ID ABB20035 standard; Protein; 118 AA.

AC ABB20035;

DT 23-JAN-2002 (first entry)

DE Protein #2034 encoded by probe for measuring heart cell gene expression.

DE Human; gene expression; heart; microarray; vascular system;

DE cardiovascular disease; hypertension; cardiac arrhythmia;

DE congenital heart disease.

OS Homo sapiens.

PN WO200157274-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00666.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

PS WPI; 2001-488899/53.

PT Single exon nucleic acid probes for analyzing gene expression in human

PT hearts -

PS Claim 15; SEQ ID No 21805; 530pp; English.

CC The present invention relates to single exon nucleic acid probes for

CC measuring human gene expression in a sample derived from human heart (see

CC ABA21535-ABA411305). The present sequence is a protein encoded by one such

CC probe. The probes may be used for predicting, measuring and displaying

CC gene expression in samples derived from the human heart via microarrays.

CC By measuring gene expression, the probes are useful for predicting,

CC diagnosing, grading, staging, monitoring and prognosing diseases of the

CC human heart and vascular system e.g. cardiovascular disease,

CC hypertension, cardiac arrhythmias and congenital heart disease.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 118 AA;

Query Match 90.8%; Score 276; DB 22; Length 118;

Best Local Similarity 88.9%; Pred. No. 2.5e-28;

Matches 56; Conservative 1; Mismatches 2; Indels 4; Gaps 1;

QY 1 KVHGLARAGKVRGQTPKVAQEKK----KKKTGRKRMQYNNRRFVNVVPTFGKKKGP 56

DB 56 KVHGLARAGKVRGQTPKVAQEKKKKKKKTGQATRMQYNNRRFVNVVPTFGKKKGP 115

QY 57 ANS 59

DB 116 ANS 118

RESULT 7

ABB20035

ID ABB20035 standard; Protein; 118 AA.

AC ABB20035;

DT 23-JAN-2002 (first entry)

DE Protein #2034 encoded by probe for measuring heart cell gene expression.

DE Human; gene expression; heart; microarray; vascular system;

DE cardiovascular disease; hypertension; cardiac arrhythmia;

DE congenital heart disease.

OS Homo sapiens.

PN WO200157274-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00666.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

PS WPI; 2001-488899/53.

PT Single exon nucleic acid probes for analyzing gene expression in human

PT hearts -

PS Claim 15; SEQ ID No 21805; 530pp; English.

CC The present invention relates to single exon nucleic acid probes for

CC measuring human gene expression in a sample derived from human heart (see

CC ABA21535-ABA411305). The present sequence is a protein encoded by one such

CC probe. The probes may be used for predicting, measuring and displaying

CC gene expression in samples derived from the human heart via microarrays.

CC By measuring gene expression, the probes are useful for predicting,

CC diagnosing, grading, staging, monitoring and prognosing diseases of the

CC human heart and vascular system e.g. cardiovascular disease,

CC hypertension, cardiac arrhythmias and congenital heart disease.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 118 AA;

Query Match 90.8%; Score 276; DB 22; Length 118;

Best Local Similarity 88.9%; Pred. No. 2.5e-28;

Matches 56; Conservative 1; Mismatches 2; Indels 4; Gaps 1;

QY 1 KVHGLARAGKVRGQTPKVAQEKK----KKKTGRKRMQYNNRRFVNVVPTFGKKKGP 56

DB 56 KVHGLARAGKVRGQTPKVAQEKKKKKKKTGQATRMQYNNRRFVNVVPTFGKKKGP 115

QY 57 ANS 59

DB 116 ANS 118

RESULT 7

ABB20035

ID ABB20035 standard; Protein; 118 AA.

AC ABB20035;

DT 23-JAN-2002 (first entry)

DE Protein #2034 encoded by probe for measuring heart cell gene expression.

DE Human; gene expression; heart; microarray; vascular system;

DE cardiovascular disease; hypertension; cardiac arrhythmia;

DE congenital heart disease.

OS Homo sapiens.

PN WO200157274-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00666.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

PS WPI; 2001-488899/53.

PT Single exon nucleic acid probes for analyzing gene expression in human

PT hearts -

PS Claim 15; SEQ ID No 21805; 530pp; English.

CC The present invention relates to single exon nucleic acid probes for

CC measuring human gene expression in a sample derived from human heart (see

CC ABA21535-ABA411305). The present sequence is a protein encoded by one such

CC probe. The probes may be used for predicting, measuring and displaying

CC gene expression in samples derived from the human heart via microarrays.

CC By measuring gene expression, the probes are useful for predicting,

CC diagnosing, grading, staging, monitoring and prognosing diseases of the

CC human heart and vascular system e.g. cardiovascular disease,

CC hypertension, cardiac arrhythmias and congenital heart disease.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 118 AA;

Query Match 90.8%; Score 276; DB 22; Length 118;

Best Local Similarity 88.9%; Pred. No. 2.5e-28;

Matches 56; Conservative 1; Mismatches 2; Indels 4; Gaps 1;

QY 1 KVHGLARAGKVRGQTPKVAQEKK----KKKTGRKRMQYNNRRFVNVVPTFGKKKGP 56

DB 56 KVHGLARAGKVRGQTPKVAQEKKKKKKKTGQATRMQYNNRRFVNVVPTFGKKKGP 115

QY 57 ANS 59

DB 116 ANS 118

RESULT 7

ABB20035

ID ABB20035 standard; Protein; 118 AA.

AC ABB20035;

DT 23-JAN-2002 (first entry)

DE Protein #2034 encoded by probe for measuring heart cell gene expression.

DE Human; gene expression; heart; microarray; vascular system;

DE cardiovascular disease; hypertension; cardiac arrhythmia;

DE congenital heart disease.

OS Homo sapiens.

PN WO200157274-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00666.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

PS WPI; 2001-488899/53.

PT Single exon nucleic acid probes for analyzing gene expression in human

PT hearts -

PS Claim 15; SEQ ID No 21805; 530pp; English.

CC The present invention relates to single exon nucleic acid probes for

CC measuring human gene expression in a sample derived from human heart (see

CC ABA21535-ABA411305). The present sequence is a protein encoded by one such

CC probe. The probes may be used for predicting, measuring and displaying

CC gene expression in samples derived from the human heart via microarrays.

CC By measuring gene expression, the probes are useful for predicting,

CC diagnosing, grading, staging, monitoring and prognosing diseases of the

CC human heart and vascular system e.g. cardiovascular disease,

CC hypertension, cardiac arrhythmias and congenital heart disease.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 118 AA;

Query Match 90.8%; Score 276; DB 22; Length 118;

Best Local Similarity 88.9%; Pred. No. 2.5e-28;

Matches 56; Conservative 1; Mismatches 2; Indels 4; Gaps 1;

QY 1 KVHGLARAGKVRGQTPKVAQEKK----KKKTGRKRMQYNNRRFVNVVPTFGKKKGP 56

DB 56 KVHGLARAGKVRGQTPKVAQEKKKKKKKTGQATRMQYNNRRFVNVVPTFGKKKGP 115

QY 57 ANS 59

DB 116 ANS 118

RESULT 7

ABB20035

ID ABB20035 standard; Protein; 118 AA.

AC ABB20035;

DT 23-JAN-2002 (first entry)

DE Protein #2034 encoded by probe for measuring heart cell gene expression.

DE Human; gene expression; heart; microarray; vascular system;

DE cardiovascular disease; hypertension; cardiac arrhythmia;

DE congenital heart disease.

OS Homo sapiens.

PN WO200157274-A2.

PD 09-AUG-2001.

PF

```
Matches 56; Conservative 1; Mismatches 2; Indels 4; Gaps 1;
QY 1 KYHGSLARAGKVGOTPRKVAKEKK-----KKKTGRAKRMQVRRFVNVVPTFGKKKGPN 56
   |||||
DB 56 KYHGSLARAGKVGOTPRKVAKEKKKKKKKTGQATRRMQVRRFVNVVPTFGKKKGPN 115
QY 57 ANS 59
   |||
DB 116 ANS 118

RESULT 8
AAM55415
ID AAM55415 standard; Protein; 118 AA.
XX
AC AAM55415;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 27520.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
PS Example 4; SEQ ID NO: 27520; 650bp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.
XX
SQ Sequence 118 AA;

Query Match 90.8%; Score 276; DB 22; Length 118;
Best Local Similarity 88.9%; Pred. No. 2.5e-28;
Matches 56; Conservative 1; Mismatches 2; Indels 4; Gaps 1;
```

```
RESULT 9
AAM67811
ID AAM67811 standard; Protein; 118 AA.
XX
AC AAM67811;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 28117.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
PS Example 4; SEQ ID NO: 28117; 658bp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention.
XX
SQ Sequence 118 AA;

Query Match 90.8%; Score 276; DB 22; Length 118;
Best Local Similarity 88.9%; Pred. No. 2.5e-28;
Matches 56; Conservative 1; Mismatches 2; Indels 4; Gaps 1;
QY 1 KYHGSLARAGKVGOTPRKVAKEKK-----KKKTGRAKRMQVRRFVNVVPTFGKKKGPN 56
   |||||
DB 56 KYHGSLARAGKVGOTPRKVAKEKKKKKKKTGQATRRMQVRRFVNVVPTFGKKKGPN 115
QY 57 ANS 59
   |||
DB 116 ANS 118

RESULT 10
AAM15620
ID AAM15620 standard; Protein; 118 AA.
XX
AC AAM15620;
XX
DT 12-OCT-2001 (first entry)
XX
DE Peptide #2054 encoded by probe for measuring cervical gene expression.
```

```
XX KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX KW cervical cancer.
XX OS Homo sapiens.
XX PN WO200157278-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00663.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX PS WPI; 2001-488901/53.
XX DR
XX XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX PS Claim 27; SEQ ID No 20446; 487pp; English.
XX CC The present invention relates to human single exon nucleic acid probes
CC (SENP: see Aa110068-AA128459). The present sequence is a peptide encoded
CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC can be used to produce a single exon microarray, which can be used for
CC measuring human gene expression in a sample derived from human cervical
CC epithelial cells. By measuring gene expression, the probes are therefore
CC useful in grading and/or staging of diseases of the cervix, notably
CC cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX XX
XX SQ Sequence 118 AA;
Query Match 90.8%; Score 276; DB 22; Length 118;
Best Local Similarity 88.9%; Pred. No. 2.5e-28;
Matches 56; Conservative 1; Mismatches 2; Indels 4; Gaps 1;
Qy 1 KVHGLARAGKVRGQTPKVAQEKK----KKKTGRAKRMQYNRRFVNVPVTFGKKKGP 56
Db 56 KVHGLARAGKVRGQTPKVAQEKKKKKKTKQATRRMQYNRRFVNVPVTFGKKKGP 115
Qy 57 ANS 59
Db 116 ANS 118
RESULT 11
AAM28118
ID AAM28118 standard; Protein; 118 AA.
XX AC AAM28118;
XX DT 17-OCT-2001 (first entry)
XX DE Peptide #2155 encoded by probe for measuring placental gene expression.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder.
XX OS Homo sapiens.
XX XX
```

```
PN WO200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00663.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX PS WPI; 2001-488897/53.
XX DR
XX XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX PS Claim 27; SEQ ID No 28387; 654pp; English.
XX CC The present invention relates to single exon nucleic acid probes (SENP:
CC see Aa131315-AA157546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders.
XX XX
XX SQ Sequence 118 AA;
Query Match 90.8%; Score 276; DB 22; Length 118;
Best Local Similarity 88.9%; Pred. No. 2.5e-28;
Matches 56; Conservative 1; Mismatches 2; Indels 4; Gaps 1;
Qy 1 KVHGLARAGKVRGQTPKVAQEKK----KKKTGRAKRMQYNRRFVNVPVTFGKKKGP 56
Db 56 KVHGLARAGKVRGQTPKVAQEKKKKKKTKQATRRMQYNRRFVNVPVTFGKKKGP 115
Qy 57 ANS 59
Db 116 ANS 118
RESULT 12
AAM03364
ID AAM03364 standard; Protein; 118 AA.
XX AC AAM03364;
XX DT 09-OCT-2001 (first entry)
XX DE Peptide #2046 encoded by probe for measuring breast gene expression.
XX KW Probe; human; breast disease; breast cancer; development disorder;
XX KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX OS Homo sapiens.
XX PN WO200157270-A2.
XX PD 09-AUG-2001.
XX PF 29-JAN-2001; 2001WO-US00661.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
```

PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-476286/51.  
 DR  
 XX Novel single exon nucleic acid probe used to measuring gene expression  
 PT in a human breast -  
 XX  
 PS Claim 27; SEQ ID No 12104; 322pp; English.  
 XX  
 CC The present invention relates to novel single exon nucleic acid probes  
 CC (see A100010-A110067). The present sequence is a peptide encoded by one  
 CC such probe. The probes are useful for measuring human gene expression in  
 CC a human breast sample, where the probe hybridizes at high stringency to a  
 CC nucleic acid expressed in the human breast. The probes are useful for  
 CC predicting, diagnosing, grading, staging, monitoring and prognosing  
 CC diseases of the human breast, particularly those diseases with polygenic  
 CC aetiology. The diseases include: breast cancer, disorders of development,  
 CC inflammatory diseases of the breast, fibrocystic changes, proliferative  
 CC breast disease and non-carcinoma tumours.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 118 AA:  
 Query Match 90.8%; Score 276; DB 22; Length 118;  
 Best Local Similarity 88.9%; Pred. No. 2.5e-28;  
 Matches 56; Conservative 1; Mismatches 2; Indels 4; Gaps 1;  
 QY 1 KVGSLARAGKVGQTPVAKOEKK---KKKTGRAKRMQVNRFPVNVPTFGKKKGN 56  
 DB 56 KVGSLARAGKVGQTPVAKOEKKKKKKKTGTATRMQVNRFPVNVPTFGKKKGN 115  
 QY 57 ANS 59  
 DB 116 ANS 118  
 RESULT 13  
 ID ABG37354  
 AC ABG37354; Peptide; 118 AA.  
 XX  
 DT 19-AUG-2002 (first entry)  
 XX  
 DE Human peptide encoded by genome-derived single exon probe SEQ ID 27019.  
 XX  
 KW Human; single exon probe; asthma; lung cancer; COPD; ILD;  
 KW chronic obstructive pulmonary disease; interstitial lung disease;  
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
 KW Heremansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
 KW primary ciliary dyskinesia; pulmonary hypertension;  
 KW hyaline membrane disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200186003-A2.  
 XX  
 PD 15-NOV-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00665.  
 XX  
 XX 04-FEB-2000; 2000US-180312P.  
 XX 26-MAY-2000; 2000US-207456P.  
 PR

PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-234687P.  
 PR 27-SEP-2000; 2000US-236359P.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2002-114183/15.  
 DR  
 XX Spatially-addressable set of single exon nucleic acid probes, used to  
 PT measure gene expression in human lung samples -  
 XX  
 PS Claim 27; SEQ ID No 27019; 634pp; English.  
 XX  
 CC The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human lung comprising single exon nucleic acid probes having one of  
 CC 12614 nucleic acid sequences mentioned in the specification, or their  
 CC complements or the 12387 open reading frames derived from the 12614  
 CC probes. Also included are a microarray comprising the novel set of  
 CC probes; the novel set of probes which hybridise at high stringency to a  
 CC nucleic acid expressed in the human lung; measuring gene expression in a  
 CC sample derived from human lung, comprising (a) contacting the array with  
 CC a collection of detectably labeled nucleic acids derived from human lung  
 CC mRNA, and (b) measuring the label detectably bound to each probe of  
 CC the array; identifying exons in a eukaryotic genome, comprising  
 CC (a) algorithmically predicting at least one exon from genomic sequences  
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably  
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
 CC having a fragment identical to the predicted exon, the probe is included  
 CC in the above mentioned microarray; assigning exons to a single gene,  
 CC comprising (a) identifying exons from genomic sequence by the method  
 CC above and (b) measuring the expression of each of the exons in several  
 CC tissues and/or cell types using hybridisation to a single exon  
 CC microarray having a probe with the exon, where a common pattern of  
 CC expression of the exons in the tissues and/or cell types indicates that  
 CC the exons should be assigned to a single gene; a peptide comprising one  
 CC of 12011 sequences, mentioned in the specification, or encoded by the  
 CC probes/open reading frames (ORF). The probes are used for gene  
 CC expression analysis, and for identifying exons in a gene, particularly  
 CC using human lung derived mRNA and for the study of lung diseases  
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease  
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,  
 CC Niemann-Pick disease, Heremansky-Pudlak syndrome, sarcoidosis, pulmonary  
 CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,  
 CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic  
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension  
 CC and hyaline membrane disease. The present sequence is a peptide/protein  
 CC encoded by a single exon probe of the invention.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 118 AA:  
 Query Match 90.8%; Score 276; DB 23; Length 118;  
 Best Local Similarity 88.9%; Pred. No. 2.5e-28;  
 Matches 56; Conservative 1; Mismatches 2; Indels 4; Gaps 1;  
 QY 1 KVGSLARAGKVGQTPVAKOEKK---KKKTGRAKRMQVNRFPVNVPTFGKKKGN 56  
 DB 56 KVGSLARAGKVGQTPVAKOEKKKKKKKTGTATRMQVNRFPVNVPTFGKKKGN 115  
 QY 57 ANS 59  
 DB 116 ANS 118

```

RESULT 14
ABB70963
ID   ABB70963 standard; Protein; 132 AA.
XX
AC   ABB70963;
XX
DT   26-MAR-2002    (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 39681.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PP   11-JUL-2000; 2000US-061415O.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
PT   WPI; 2001-656860/75.
DR   N-PSTB; ABLI5066.
XX
CC   New isolated nucleic acid detection reagent for detecting 1000 or more
CT   genes from Drosophila and for elucidating cell signalling and cell-cell
CS   interactions .
XX
PS   Disclosure; SEQ ID NO 39681; 2lpp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CA   capable of detecting 1000 or more genes from drosophila. The invention is
CO   useful in developmental biology and in elucidating cell signalling and
CW   cell-cell interactions in higher eukaryotes for the development of
CG   insecticides, therapeutics and pharmaceutical drugs. The invention
CH   discloses genomic DNA sequences (ABLI16176-ABLI30511), expressed DNA
CI   sequences (ABLI01840-ABLI16175) and the encoded proteins
CX   (ABBS7737-ABBS72072).
CC   The sequence data for this patent did not form part of the printed
CF   specification, but was obtained in electronic format directly from WIPO
CG   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence      132 AA;
Query Match              86.5%; Score 263; DB 22; Length 132;
Best Local Similarity    84.7%; Pred.No. 1.4e-26;
Matches 50; Conservative 5; Mismatches 4; Indels 0; Gaps 0
QY       1 KVFHSLRAGKVRGQTPKVAKOEKKKKKTGRAKRMRQMNRFRVNVPFVGKKKGPNANS 59
        ||| | | | | | | | | | | | | | | | | | | | | | : | | | | |
DB       73 KVHGSLRAGKVGGTTPKVEKOEKKKKKTGRAKRIQYNRRFNVFPQGFGERRGPANNS 131
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 15
AAG22753
ID   AAG22753 standard; Protein; 62 AA.
XX
AC   AAG22753;
XX
DT   17-OCT-2000    (first entry)
XX
DE   Zea mays protein fragment SEQ ID NO: 25803.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence; corn.
```



```
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143452.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151086.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.

PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0158294.
PR 13-OCT-1999; 99US-0158295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 78.6%; Score 239; DB 21; Length 62;
Best Local Similarity 79.3%; Pred. No. 8,4e-24;
Matches 46; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Qy 1 KKHGSLARAGKVGOTPPKAKOEKKKKKTGAKRRMQYNNRRFNVVPTFGKKKGNAN 58
Db 3 KKHGSLARAGKVGOTPPKAKODKKKKPRGAKRKQYNNRRFVTAVVFGKKRGNSS 60

Search completed: December 3, 2002, 14:31:59
Job time : 36 secs
```



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2002, 14:30:48 ; Search time 20 Seconds  
(without alignments)  
283.597 Million cell updates/sec

Title: US-09-424-815E-1

Perfect score: 304

Sequence: 1 KVHGSILARAGKVRGQTPKVA.....RRFNVVPTFGKKKGPNNAS 59

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries.

Database :

1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	304	100.0	133	1 A47416	ubiquitin-like pro
2	304	100.0	133	2 JCI278	ubiquitin-like pro
3	304	100.0	133	2 T48346	ribosomal protein
4	231	76.0	62	2 H85342	RIBOSOMAL PROTEIN
5	231	76.0	68	2 F84580	40S ribosomal prot
6	227	74.7	130	2 T15642	hypothetical prote
7	188.5	62.0	58	2 A71604	ribosomal protein
8	176	57.9	61	2 T39834	ribosomal protein
9	175.5	57.7	63	1 S67074	ribosomal protein
10	107	35.2	229	2 S36383	ribosomal protein
11	104.5	34.4	52	2 G90244	SEB4B protein - hu
12	91.5	30.1	50	2 C72528	15U ribosomal prot
13	66	21.7	212	2 S73991	probable ribosomal
14	64.5	21.2	454	1 A45340	ribosomal protein
15	64.5	21.2	454	1 B45340	nucleocapsid prote
16	64.5	21.2	454	1 C45340	nucleocapsid prote
17	64.5	21.2	454	1 A45396	nucleocapsid prote
18	64.5	21.2	455	1 VHIHMJ	nucleocapsid prote
19	64.5	21.2	455	1 D45340	nucleocapsid prote
20	64	21.1	225	2 AC1896	nucleocapsid prote
21	63.5	20.9	1526	2 A45605	hypothetical prote
22	63	20.6	207	2 JCI1129	mature-parasite-in
23	62.5	20.6	207	2 JCI1129	hypothetical prote
24	62.5	20.6	207	2 JCI1129	nonhistone chromos
25	60.5	19.9	186	2 B61611	high-mobility grou
26	60.5	19.9	617	2 S27389	nonhistone chromos
27	60	19.7	439	2 UQ1729	secretogranin II -
28	60	19.7	521	2 T37504	ankyrin-repeat pro
29	59.5	19.6	205	2 S34919	hypothetical prote
					ribosomal protein

30	59.5	19.6	752	2 S64750	probable ATP-depen
31	59	19.4	133	2 H87589	EF hand domain pro
32	59	19.4	304	2 S44801	PI0E9.2 protein -
33	58	19.1	186	2 S30221	nonhistone chromos
34	58	19.1	209	1 NSHUR2	nonhistone chromos
35	58	19.1	210	2 A34719	nonhistone chromos
36	58	19.1	377	2 G84857	hypothetical prote
37	58	19.1	774	2 A24057	glycophorin-bindin
38	58	19.1	894	2 T15769	hypothetical prote
39	57.5	18.9	211	2 H64216	ribosomal protein
40	57.5	18.9	340	2 T33559	probable transposa
41	57.5	18.9	340	2 A80096	transposase for in
42	57.5	18.9	340	2 A80198	transposase for in
43	57.5	18.9	340	2 A80395	transposase for in
44	57.5	18.9	340	2 A80488	transposase for in
45	57.5	18.9	340	2 AC0031	transposase for in

## ALIGNMENTS

RESULT 1  
A47416  
ubiquitin-like protein / ribosomal protein S30; cytosolic [validated] - rat  
N/Contains: ribosomal protein S30; ubiquitin-like protein  
C/Species: Rattus norvegicus (Norway rat)  
C/Date: 21-Jul-2000 #sequence revision 21-Jul-2000 #text\_change 21-Jul-2000  
C/Accession: A47416; B47416; S18101  
R/Olivera, J.; Wool, I.G.  
J. Biol. Chem. 268, 17967-17974, 1993  
A/Title: The carboxyl extension of a ubiquitin-like protein is rat ribosomal protein S30.  
A/Reference number: A47416; MUID:93352612; PMID:8394356  
A/Accession: A47416  
A/Molecule type: mRNA  
A/Residues: 1-133 <OLV>  
A/Cross-References: EMBL:X62671; NID:9407165; PIDN:CAA44545.1; PID:957566  
A/Accession: B47416  
A/Molecule type: protein  
A/Residues: 75-92 <OLV>  
A/Note: The proteins are designated as ubiquitin-like protein and ribosomal protein S30  
C/Suprafamily: ubiquitin-like protein / rat ribosomal protein S30; ubiquitin homology  
C/Keywords: protein biosynthesis; ribosome  
F.1-74/Product: ubiquitin-like protein #status predicted <UBI>  
F.1-74/Domains: ubiquitin homology <UBH>  
F.75-133/Product: ribosomal protein S30 #status experimental <RIB>

Query Match 100.0%; Score 304; DB 1; Length 133;  
Best Local Similarity 100.0%; Pred. No. 2.8e-25;  
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KVHGSILARAGKVRGQTPKVAKEKKKTKGRKRMQVRRVNVPTFGKKKGPNNAS 59  
Db 75 KVHGSILARAGKVRGQTPKVAKEKKKTKGRKRRQVRRVNVPTFGKKKGPNNAS 133

RESULT 2  
JCI278  
ubiquitin-like protein / ribosomal protein S30; cytosolic - human  
N/Alternate names: fau protein  
C/Species: Homo sapiens (man)  
C/Date: 30-Sep-1993 #sequence revision 30-Sep-1993 #text\_change 24-Sep-1999  
A/Title: Genomic structure and expression of the human fau gene: Encoding the ribosomal  
A/Reference number: JCI278; MUID:92412144; PMID:1326960  
A/Accession: JCI278  
A/Molecule type: DNA  
A/Residues: 1-133 <KAS>  
A/Cross-References: EMBL:X65921; NID:g31304; PIDN:CAA46714.1; PID:g31305  
R/Michels, L.; Van der Raaij-van der Wal, E.; Van Hasselt, F.; Kas, K.; Merregaert, J.  
Oncogene 8, 2537-2546, 1993  
A/Title: fau cDNA encodes a ubiquitin-like-S30 fusion protein and is expressed as an anti

A:Reference number: I37387; MUID:93368957; PMID:8395683  
A:Accession: I37387  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-133 <RES>  
A:Cross-references: EMBL:X65923; NID:g311302; PIDN:CAA46716.1; PID:g31303  
R:Vladimirov, S.N.; Ivanov, A.V.; Karpova, G.G.; Musolyamov, A.K.; Egorov, T.A.; Thiede, Eur. J. Biochem. 239, 144-149, 1996  
A:Title: Characterization of the human small-ribosomal-subunit proteins by N-terminal an  
A:Reference number: S68911; MUID:96305378; PMID:8706699  
A:Accession: S68911  
A:Molecule type: protein  
A:Residues: 75-99 <VLA>  
C:Genetics:  
A:Gene: fau  
A:Introns: 25/3; 74/1; 92/3  
C:Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology  
C:Keywords: protein biosynthesis; ribosome  
F:1-74/Domain: ubiquitin homology <UBH>  
F:75-133/Product: ribosomal protein S30, cytosolic #status experimental <MAT>  
  
Query Match 100.0%; Score 304; DB 2; Length 133;  
Best Local Similarity 100.0%; Pred. No. 2.8e-25;  
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 KVHGLARAGKVRGQTPKVAQKQKKKTGAKRMMQYNRRFVNVVPTFGKKKGPNNAS 59  
Db 75 KVHGLARAGKVRGQTPKVAQKQKKKTGAKRMMQYNRRFVNVVPTFGKKKGPNNAS 133  
  
RESULT 3  
148346  
ribosomal protein fau - mouse  
N:Alternate names: gene fau protein; monoclonal nonspecific suppressor factor beta  
C:Species: Mus musculus (house mouse)  
C:Date: 12-Aug-1996 #sequence revision 12-Aug-1996 #text\_change 24-Sep-1999  
R:Accession: I48346; AS6532; I59368; S21452  
R:Michaels, L.; Van der Raaij-MacLeod, E.; Van Hasselt, F.; Kas, K.; Merregaert, J.  
Oncogene 8, 2537-2546, 1993  
A:Title: fau cDNA encodes a ubiquitin-like-S30 fusion protein and is expressed as an an  
A:Reference number: I37387; MUID:93368957; PMID:8395683  
A:Accession: I48346  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-133 <RES>  
A:Cross-references: EMBL:X65922; NID:g50949; PIDN:CAA46715.1; PID:g50950  
R:Castells, D.; Poirier, C.; Guenet, J.L.; Merregaert, J.  
Genomics 25, 291-294, 1995  
A:Title: The mouse Fau gene: genomic structure, chromosomal localization, and characteri  
A:Reference number: AS6532; MUID:95293388; PMID:7774934  
A:Accession: AS6532  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-133 <CAS>  
A:Cross-references: GB:I33715; NID:g497610; PIDN:AAA91564.1; PID:g497611  
A:Note: authors translated the codon GTT for residue 119 as Arg, and GTC for residue 120  
R:Nakamura, M.; Xavier, R.M.; Tsunematsu, T.; Tanigawa, Y.  
Proc. Natl. Acad. Sci. U.S.A. 92, 3463-3467, 1995  
A:Title: Molecular cloning and characterization of a cDNA encoding monoclonal nonspecifi  
A:Reference number: I59368; MUID:95241522; PMID:7724584  
A:Accession: I59368  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-133 <RE2>  
A:Cross-references: GB:D26610; NID:g1060926; PIDN:BAA05655.1; PID:g1060927  
C:Genetics:  
A:Gene: fau  
A:Introns: 25/3; 74/1; 92/3  
C:Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology  
F:1-74/Domain: ubiquitin homology <UBH>  
  
Query Match 100.0%; Score 304; DB 2; Length 133;  
Best Local Similarity 100.0%; Pred. No. 2.8e-25;

Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 KVHGLARAGKVRGQTPKVAQKQKKKTGAKRMMQYNRRFVNVVPTFGKKKGPNNAS 59  
Db 75 KVHGLARAGKVRGQTPKVAQKQKKKTGAKRMMQYNRRFVNVVPTFGKKKGPNNAS 133  
  
RESULT 4  
H85342  
RIBOSOMAL PROTEIN S30 homolog [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001  
C:Accession: H85342  
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spr  
Nature 402, 769-777, 1999  
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
A:Reference number: A85001; MUID:20083488; PMID:10617198  
A:Accession: H85342  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-62 <STO>  
A:Cross-references: GB:NC\_001268; NID:g7269837; PIDN:CAB79697.1; GSPDB:GN00140  
C:Genetics:  
A:Gene: At4g29390  
A:Map position: 4  
  
Query Match 76.0%; Score 231; DB 2; Length 62;  
Best Local Similarity 75.9%; Pred. No. 8e-18;  
Matches 44; Conservative 7; Mismatches 7; Indels 0; Gaps 0;  
  
Qy 1 KVHGLARAGKVRGQTPKVAQKQKKKTGAKRMMQYNRRFVNVVPTFGKKKGPNNAN 58  
Db 3 KVHGLARAGKVRGQTPKVAQKQKKKPRGAKHKLQHNRRFVTAVVVGFGKKRGPNSS 60  
  
RESULT 5  
F84580  
40S ribosomal protein S30 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 17-May-2002  
C:Accession: F84580  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, I.  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487; PMID:10617197  
A:Accession: F84580  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-68 <STO>  
A:Cross-references: GB:AE002093; NID:g3687243; PIDN:AAC62141.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g19750  
A:Map position: 2  
C:Superfamily: yeast ribosomal protein S30.e  
  
Query Match 76.0%; Score 231; DB 2; Length 68;  
Best Local Similarity 75.9%; Pred. No. 8.7e-18;  
Matches 44; Conservative 7; Mismatches 7; Indels 0; Gaps 0;  
  
Qy 1 KVHGLARAGKVRGQTPKVAQKQKKKTGAKRMMQYNRRFVNVVPTFGKKKGPNNAN 58  
Db 9 KVHGLARAGKVRGQTPKVAQKQKKKPRGAKHKLQHNRRFVTAVVVGFGKKRGPNSS 66  
  
RESULT 6  
T15642  
hypotheical protein C26F1.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 24-Nov-1999  
C:Accession: T15642  
R:Geisler, C.; Stellyes, L.; Bradshaw, H.



Qy	3	HGSLARAGKVRGOTPKV-AKOEKKKKKTGRAKRRMQYNRRFVN	45
		:	
Dd	4	HGSLTRAGKVRKQTPLPAKQ--KKNYPPRLKNRLKYQVRIEKV	45
		:	
 RESULT 13 S73991			
ribosomal protein L4 - Mycoplasma pneumoniae (strain ATCC			
N;Alternate names: hypothetical protein VxpSP7_orf212			
C;Species: Mycoplasma pneumoniae			
A;Variety: ATCC 29342			
C;Date: 27-Feb-1997 #sequence_revision 25-Apr-1997 #text_			
C;Accession: S73991			
R;Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; L			
Nucleic Acids Res. 24, 4420-4449, 1996			
A;Title: Complete sequence analysis of the genome of the			
A;Reference number: S73327; MUID:97105885; PMID:8948633			
A;Accession: S73991			
A;Status: nucleic acid sequence not shown; translation not			
A;Molecule type: DNA			
A;Residues: 1-212 <IM>			
A;Cross-references: EMBL:AE000061; GB:U00089; NID:g167433			
A;Note: the nucleotide sequence was submitted to the EMBL			
C;Genetics:			
A;Gene: rplD			
A;Genetic code: GSC3			
C;Superfamily: Escherichia coli ribosomal protein L4			
C;Keywords: protein biosynthesis; ribosome			
 Query Match            21.7%;    Score 66;    DB 2;    Length			
Best Local Similarity    37.0%;    Pred. No. 7.1;			
Matches         20;    Conservative         7;    Mismatches         21;    In			
 Qy         5   SLARAGKVRGOTEKVAKOEKKKKKTGRAKRRMQYNRRFVNVPTEGK			
		:	
Dd	51	SILTKGEVRGG---GKPYKQKHGTGARQGSTRPHFGGVIVPGP	
		:	
 RESULT 14			

A15340  
nucleocapsid protein - murine hepatitis virus (strain A59  
C/Species: murine hepatitis virus, MHV  
A/Variety: strain A59  
C/Date: 30-Sep-1993 #sequence revision 30-Sep-1993 #text\_  
C/Accession: A45340; A04023; A47310  
R/Parker, M.W.; Masters, P.S.  
Virology 179, 463-468, 1990  
A/Title: Sequence comparison of the N genes of five strain  
A/Reference number: A45340; MUID:91021052; PMID:2171216  
A/Accession: A45340  
A/Molecule type: genomic RNA  
A/Residues: 1-454 <PAR>  
A/Cross-references: GB:X5256; NID:g331826; PIDN:AAA46447  
R/Armstrong, J.; Smeekens, S.; Rottier, P.  
Nucleic Acids Res. 11, 883-891, 1983  
A/Title: Sequence of the nucleocapsid gene from murine co  
A/Reference number: A04023; MUID:83168908; PMID:6687635  
A/Accession: A04023  
A/Molecule type: genomic RNA  
A/Residues: 1-108, 'AVLLHLMGRSNVCPDIGFTILAQGPMLEPVMETALKK  
A/Cross-references: GB:X0509; GB:J02252; NID:g58965  
R/Schaad, M.C.; Baric, R.S.  
Virology 196, 190-198, 1993  
A/Title: Evidence for new transcriptional units encoded a  
A/Reference number: A47310; MUID:93362405; PMID:8395114  
A/Accession: A47310  
A/Molecule type: mRNA  
A/Residues: 301-454 <SCH>  
A/Cross-references: GB:S64884; NID:g408334; PIDN:AAB27902  
A/Note: sequence extracted from NCBI backbone (NCBIN:1365)  
C/Genetics:  
A/Gen: N







GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2002, 14:30:48 ; Search time 11 seconds  
(without alignments)  
222.464 Million cell updates/sec

Title: US-09-424-815E-1

Perfect score: 304  
Sequence: 1 KVHSLARAGKVRGQTPKVA.....RRFNVVPTFGKKKGNANS 59

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	304	100.0	59 1 RS30_HUMAN	Q05472 homo sapien
2	254	83.6	59 1 RS30_ORYLA	Q9WY0 Oryzias lat
3	231	76.0	62 1 RS30_ARATH	P49689 arabidopsis
4	188.5	62.0	58 1 RS30_PLAFA	O96269 plasmodium
5	176	57.9	61 1 RS30_SCHPO	O42952 schizosacch
6	175.5	57.7	62 1 RS30_YEAST	Q12087 saccharomyc
7	91.5	30.1	50 1 RS30_AERPE	Q9Y3C9 aetopyrum p
8	67.5	22.2	454 1 NCAL_CVMA5	P03416 murine coro
9	66	21.7	212 1 RLA_MYCPN	P75579 mycoplasma
10	64.5	21.2	454 1 NCAL_CVMA5	P18448 murine coro
11	64.5	21.2	454 1 NCAL_CVMA5	P18447 murine coro
12	64.5	21.2	454 1 NCAL_CVMA5	P18449 murine coro
13	64.5	21.2	454 1 NCAL_CVMA5	Q02915 rat coronav
14	64.5	21.2	455 1 NCAL_CVMA5	P18446 murine coro
15	64.5	21.2	455 1 NCAL_CVMA5	P03417 murine coro
16	63	20.7	399 1 CC37_DROVI	Q24740 dirosophila
17	62.5	20.6	206 1 HMG2_CHICK	P26584 gallus gall
18	60.5	19.9	617 1 SG2_MOUSE	Q03517 mus musculu
19	60.5	19.9	1362 1 BRD4_HUMAN	O60885 homo sapien
20	60	19.7	439 1 AKR_ARATH	Q05753 arabidopsis
21	60	19.7	1064 1 ISK5_HUMAN	Q9nq38 homo sapien
22	59.5	19.6	205 1 RRA_EUGAR	P27418 euglena gra
23	59.5	19.6	752 1 DR61_YEAST	P32892 saccharomyc
24	59.5	19.4	286 1 YLU2_CAEEL	P34396 caenorhabdi
25	58	19.1	208 1 HMG2_HUMAN	P26583 homo sapien
26	58	19.1	209 1 HMG2_PIG	P17741 sus scrofa
27	58	19.1	774 1 GLYB_PLAFA	P02895 plasmodium
28	57.5	18.9	211 1 RLA_MYCGE	P47398 mycoplasma
29	57.5	18.9	555 1 PERE_HUMAN	P14222 homo sapien
30	57.5	18.9	862 1 VGO1_HSV11	Q00132 icetalarid h
31	57	18.8	209 1 HMG2_MOUSE	P30681 mus musculu
32	57	18.8	209 1 HMG2_RAT	P52825 rattus norv
33	57	18.8	685 1 SNMA_DICDI	P54705 dictyosteli

34	57	18.8	785 1 MUS2_BACHD	Q9k8a0 bacillus ha
35	56.5	18.6	372 1 CIS2_BACCU	P39120 bacillus su
36	56	18.4	150 1 VPS_BPP2	P36934 bacterioph
37	56	18.4	174 1 YD5A_SCHPO	O14185 schizosacch
38	55.5	18.3	187 1 PENK_SCHPO	O28409 felis silve
39	55.5	18.3	233 1 YQAL_BACCU	P45909 bacillus su
40	55	18.1	11 1 RS30_ONCMY	P83328 oncorhynch
41	55	18.1	121 1 H2B_ASTRU	P02286 asterias ru
42	55	18.1	747 1 DIL2_HUMAN	Q911w0 homo sapien
43	55	18.1	2059 1 TEGU_HSV7J	P52362 human herpe
44	54.5	17.9	110 1 Y225_METYA	Q60284 methanococ
45	54.5	17.9	209 1 R44_MYCGA	O52333 mycoplasma

## ALIGNMENTS

RESULT 1  
ID RS30\_HUMAN STANDARD; PRT; 59 AA.  
AC Q05472; Q95261;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 40S ribosomal protein S30.  
GN PAU.  
OS Homo sapiens (Human).  
OS Mus musculus (Mouse).  
OS Rattus norvegicus (Rat).  
OS Cricetulus griseus (Chinese hamster), and  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
NCBI\_Taxid=9606, 10090, 10116, 10029, 9823;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=Human;  
RX MEDLINE=92412144; PubMed=1326960;  
RA Kas K., Michiels L., Merregaert J.;  
RT "Genomic structure and expression of the human fau gene: encoding the  
RT ribosomal protein S30 fused to a ubiquitin-like protein.";  
RT Biochem. Biophys. Res. Commun. 187:927-933(1992).  
[2]  
RP SEQUENCE FROM N.A.  
RC SPECIES=Human, and Mouse;  
RX MEDLINE=93368957; PubMed=8395683;  
RA Michiels L., Van der Raaij-MacLeod E., van Haeseelt F., Kas K.,  
RA Merregaert J.;  
RT "faa cDNA encodes a ubiquitin-like-S30 fusion protein and is  
RT expressed as an antisense sequence in the Finkel-Biskis-Reilly murine  
RT sarcoma virus.";  
RT Oncogene 8:2537-2546(1993).  
[3]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-18.  
RC SPECIES=Rat; STRAIN=Sprague-Dawley; TISSUE=liver;  
RX MEDLINE=93352612; PubMed=8394356;  
RA Olivera J., Wool I.G.;  
RT "The carboxyl extension of a ubiquitin-like protein is rat ribosomal  
RT protein S30.";  
RT J. Biol. Chem. 268:17967-17974(1993).  
[4]  
RP SEQUENCE FROM N.A.  
RC SPECIES=Mouse; STRAIN=BALB/C;  
RX MEDLINE=95241522; PubMed=7724584;  
RA Nakamura M., Xavier R.W., Tsunematsu T., Tanigawa Y.;  
RT "Molecular cloning and characterization of a cDNA encoding monoclonal  
RT non-specific suppressor factor.";  
RL Proc. Natl. Acad. Sci. U.S.A. 92:3463-3467(1995).  
[5]  
RP SEQUENCE FROM N.A.  
RC SPECIES=Mouse; STRAIN=BALB/C; TISSUE=liver;  
RX MEDLINE=95293388; PubMed=7774934;  
RA Casteels D., Poltier C., Guenet J.-L., Merregaert J.;

```
RT "The mouse Fau gene: genomic structure, chromosomal localization, and
RL characterization of two retrospseudogenes.";
RN Genomics 25:291-294 (1995).
[6]
RP SEQUENCE FROM N.A.
RC SPECIES=C. griseus;
RA Rossmann T.G., Wang Z.;
RN Submitted (DSC-1995) to the EMBL/GenBank/DBJ databases.
[7]
RP SEQUENCE FROM N.A.
RC SPECIES=Pig; TISSUE=Uterus;
RX MEDLINE=9724440; PubMed=9089280;
RA Chwetroff S., D'Andrea S.;
RT "Ubiquitin is physiologically induced by interferons in luminal
RT epithelium of porcine uterine endometrium in early pregnancy: global
RT RT-PCR cDNA in place of RNA for differential display screening.";
RL FEBS Lett. 405:148-152 (1997).
CC -|- MISCELLANEOUS: THIS RIBOSOMAL PROTEIN IS SYNTHESIZED AS A
CC C-TERMINAL EXTENSION PROTEIN (CEP) OF A UBIQUITIN-LIKE PROTEIN.
CC -|- SIMILARITY: BELONGS TO THE S30E FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ238274; CAB40969.1; ALT_INIT.
DR Ribosomal protein.
KW Ribosomal protein.
SQ SEQUENCE 59 AA; 6648 MW; 012AC1FE555B01A4 CRC64;

Query Match 100.0%; Score 304; DB 1; Length 59;
Best Local Similarity 100.0%; Pred. No. 3.2e-29;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KVHGLARAGKVRGQTPKVAQKQKKKTGAKRMRQYNRRFVNVVPTFGKKKGPNANS 59
Db 1 KVHGLARAGKVRGQTPKVAQKQKKKTGAKRMRQYNRRFVNVVPTFGKKKGPNANS 59

RESULT 2
RS30_ORYLA STANDARD; PRT; 59 AA.
ID RS30_ORYLA
AC Q9W6Y0;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 40S ribosomal protein S30.
GN FAU OR RPS30.
OS Oryzias latipes (Medaka fish) (Japanese ricefish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Atherinomorpha;
OC Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
OX NCBI_TaxID=8090;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20070552; PubMed=10603084;
RA Heinrich T., Wittbrodt J.;
```

```
RT "An in situ hybridization screen for the rapid isolation of
RL differentially expressed genes.";
RN Dev. Genes Evol. 210:28-33 (2000).
CC -|- MISCELLANEOUS: THIS RIBOSOMAL PROTEIN IS SYNTHESIZED AS A
CC C-TERMINAL EXTENSION PROTEIN (CEP) OF A UBIQUITIN-LIKE PROTEIN.
CC -|- SIMILARITY: BELONGS TO THE S30E FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ238274; CAB40969.1; ALT_INIT.
DR Ribosomal protein.
KW Ribosomal protein.
SQ SEQUENCE 59 AA; 6660 MW; E91B9C088C7772CF CRC64;

Query Match 83.6%; Score 254; DB 1; Length 59;
Best Local Similarity 83.1%; Pred. No. 2.3e-23;
Matches 49; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Qy 1 KVHGLARAGKVRGQTPKVAQKQKKKTGAKRMRQYNRRFVNVVPTFGKKKGPNANS 59
Db 1 KVHGLARAGKVRGQTPKVAQKQKKKTGAKRMRQYNRRFVNVVPTFGKKKGPNANS 59

RESULT 3
RS30_ARATH STANDARD; PRT; 62 AA.
ID RS30_ARATH
AC P49689; O82203; Q9M0E4;
DT 01-FEB-1996 (Rel. 33, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 40S ribosomal protein S30.
GN (RPS30A OR AT2G19750 OR P6F22.22) AND
GN (RPS30B OR AT4G29390 OR F1AI3.210) AND
GN (RPS30C OR AT5G56670 OR MIK19.12).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsiis.
OX NCBI_TaxID=3702;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Lebas M., Regad F., Lescure B.;
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A. (RPS30A).
RC STRAIN=cv. Columbia;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,
RA Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,
RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana.";
RL Nature 402:761-768 (1999).
[3]
RP SEQUENCE FROM N.A. (RPS30B).
RC STRAIN=cv. Columbia;
RX MEDLINE=20083488; PubMed=10617198;
RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terry N.,
RA Harris B., Ansoerge W., Brandt P., Grivell L.A., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
RA Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidheini T.,
```

RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,  
 RA Vos P., Hohelsel J., Zimmermann W., Wedler H., Ridley P.,  
 RA Langham S.-A., McCallagh B., Bilham L., Robben J.,  
 RA Van der Schueren J., Grymoprez B., Chuang Y.-J., Vandenhussche F.,  
 RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,  
 RA Weizenecker T., Bothe G., Rameberger U., Hilbert H., Braun M.,  
 RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,  
 RA Mooljman P., Klein Lankhorst R., Rose M., Hauf J., Koelster P.,  
 RA Bernieris S., Hempel S., Feldpausch M., Lambirth S., Van den Daele H.,  
 RA De Keyser A., Buysheert C., Gielen J., Villarroel R., De Clercq R.,  
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,  
 RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McIay K., Mayes R.,  
 RA Petrett A., Rajandream M.A., Lyne M., Benes V., Reckmann S.,  
 RA Borrova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,  
 RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Farman B., Granderath K., Danner D., Herzi A.,  
 RA Neumann S., Argitrou A., Vitale D., Liguori R., Piravanni E.,  
 RA Maesene O., Ougley F., Clabaud G., Muendlein A., Felber R.,  
 RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,  
 RA Chetdor F., Cooke R., Berger C., Monfort A., Casacuberta E.,  
 RA Gibbons T., Weber N., Vanderbol M., Barques M., Terol J., Torres A.,  
 RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,  
 RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,  
 RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stockert S.,  
 RA Zaccaria P., Bevan M., Wilson R.K., de la Baetide M., Habermann K.,  
 RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,  
 RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,  
 RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,  
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,  
 RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,  
 RA Nelson J., Speth J., Ryan E., Andrews S., Geisel C., Layman D.,  
 RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,  
 RA Antonou B., Zidan M., Strong C., Sun H., Lamar B., Yordan C.,  
 RA Ma P., Zhong J., Preston R., Val D., Shekher M., Marero A., Shah R.,  
 RA Swaby J.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,  
 RA Gravat S., Shohdy N., Hasegawa A., Hamed A., Lodhi M., Johnson A.,  
 RA Chen E., Marra M., Martensen R., McCombie W.R.,  
 RA "Sequence and analysis of chromosome 4 of the plant Arabidopsis  
 RT thaliana."  
 RT Nature 402:769-777(1999).  
 RL [4]  
 RN SEQUENCE FROM N.A. (RPS30C).  
 RP STRAIN=cv. Columbia;  
 RC MEDLINE=98403884; PubMed=9734815;  
 RX MEDLINE=98403884; PubMed=9734815;  
 RA Kotani H., Nakamura Y., Sato S., Asamizu E., Kaneko T., Miyajima N.,  
 RA Tabata S.,  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. VI.  
 RT Sequence features of the regions of 1,367,185 bp covered by 19  
 RT physically assigned P1 and TAC clones.";  
 RL DNA Res. 5:203-216(1998).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.,  
 RT "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the  
 RT SSP consortium (Salk/Stanford/PECC)."  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 CC -I- SIMILARITY: BELONGS TO THE S30E FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z26869; CAA81482.1; -;  
 DR EMBL; AC006169; AAC62141.2; -;  
 DR EMBL; AL161574; CAB79697.1; -;  
 DR EMBL; AL096692; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; AB013392; BAB09885.1; -;

DR EMBL; AY052341; AAK96533.1; -;  
 DR EMBL; AY061910; AAL31237.1; -;  
 KW Ribosomal protein, Multigene family.  
 FT CONFLICT 17 19 OTP -> RHQ (IN REF. 1).  
 SQ SEQUENCE 62 AA; 6887 MW; 95D8F3EB72F53F33 CRC64;  
 Query Match 76.0%; Score 231; DB 1; Length 62;  
 Best Local Similarity 75.9%; Pred. No. 1.2e-20;  
 Matches 44; Conservative 7; Mismatches 7; Indels 0; Gaps 0;  
 QY 1 KYHGSILARAGKVRGQTPKVAKOEKKKKTKGRAKRQNVRRPNNVPTGKKKGPAN 58  
 DB 3 KYHGSILARAGKVRGQTPKVAKODKKKKRGRHAKRLQHNRRFVAVVGFCKKGPNS 60  
 RESULT 4  
 RS30\_PLAFA STANDARD; PRT; 58 AA.  
 ID RS30\_PLAFA  
 AC 096269;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE 40S ribosomal protein S30.  
 GN RPS30 OR PFB0885W.  
 OS Plasmodium falciparum.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxId=5833;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99021743; PubMed=9804551;  
 RA Gardner M.J., Tetteijn H., Carucci D.J., Cummings L.M., Aravind L.,  
 RA Koonin E.V., Shallow S., Mason T., Yu K., Fujii C., Pederson J.,  
 RA Shen K., Jing J., Asencio C., Lai Z., Schwartz D.C., Petrea M.,  
 RA Salberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,  
 RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.,  
 RT "Chromosome 2 sequence of the human malaria parasite Plasmodium  
 RT falciparum."  
 RL Science 282:1126-1132(1998).  
 CC -I- SIMILARITY: BELONGS TO THE S30E FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE001422; AAC71966.1; -;  
 DR EMBL; AE001422; AAC71966.1; -;  
 KW Ribosomal protein.  
 SQ SEQUENCE 58 AA; 6488 MW; 525675AE74D04ESF CRC64;  
 Query Match 62.0%; Score 188.5; DB 1; Length 58;  
 Best Local Similarity 64.4%; Pred. No. 1.1e-15;  
 Matches 38; Conservative 8; Mismatches 10; Indels 3; Gaps 1;  
 QY 1 KYHGSILARAGKVRGQTPKVAKOEKKKKTKGRAKRQNVRRPNNVPTGKKKGPAN 59  
 DB 3 KYHGSILARAGKVRGQTPKVAKODKKKKRGRHAKRLQHNRRFVAVVGFCKKGPNS 60  
 RESULT 5  
 RS30\_SCHPO STANDARD; PRT; 61 AA.  
 ID RS30\_SCHPO  
 AC 042952; O14314;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE 40S ribosomal protein S30.  
 GN RPS30A OR RPS30 OR SPAC19B12.04 OR (RPS30B OR SPAC19G7.03C).  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;

(RPS30A OR YLR287BC OR L8003.23) AND (RPS30B OR YOR182C).  
Saccharomyces cerevisiae (Baker's Yeast).  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
NCBI\_TaxID=4932;  
[1]\_TaxID=4932;  
SEQUENCE FROM N.A., AND SEQUENCE (RPS30A).  
STRAIN=YRBI41;  
MEDLINE=96278780; PubMed=8662789;  
Baker R.T., Williams N.A., Wettenthal R.E.H.;  
"The yeast homolog of mammalian ribosomal protein S30 is expressed  
from a duplicated gene without a ubiquitin-like protein fusion  
sequence. Evolutionary implications.";  
RL J. Biol. Chem. 271:13549-13555(1996).  
[2]  
SEQUENCE FROM N.A. (RPS30A).  
R Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,  
Pavello A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,  
Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,  
R Johnston L., Langston Y., Latreille P., Le T., Mardis E., Menszes S.,  
Miller N., Nhan M., Pauley A., Peluso D., Rifken L., Riles L.,  
Taich A., Trevasik E., Vignati D., Wilcox L., Wohlman P., Vaudin M.,  
Wilson R., Waterston R.;  
Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
[3]  
SEQUENCE FROM N.A. (RPS30B).  
R Hughes B., Pohl T.M.;  
Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
CC - MASS SPECTROMETRY: MW=6987; MW ERR=3.4; METHOD=MALDI.  
CC - MISCELLANEOUS: THERE ARE TWO GENES FOR S30 IN YEAST.  
CC - SIMILARITY: BELONGS TO THE S30E FAMILY OF RIBOSOMAL PROTEINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U48700; AAC49317.1; -  
DR EMBL; U48699; AAC49316.1; -  
DR EMBL; U83406; AAB41050.1; -  
DR EMBL; U83407; AAB41051.1; -  
DR EMBL; U17243; AAB67333.1; -  
DR EMBL; U75090; CAA99391.1; -  
DR SGD; S0004278; RPS30A.  
DR SGD; S0005708; RPS30B.  
KW Ribosomal protein; Multigene family.  
FT INIT MET 0  
SQ SEQUENCE 62 AA; 6987 MW; 65EADAED8D9FC586 CRC64;  
  
Query Match 57.7%; Score 175.5; DB 1; Length 62;  
Best Local Similarity 62.7%; Pred. No. 3.8e-14;  
Matches 37; Conservative 6; Mismatches 13; Indels 3; Gaps 1;  
  
QY 1 KVHGLARAGKVGOTPKVAKOEKKKKTKGRAKRMQYNRRFVNVPVTFGKKK---GPN 56  
Db 2 KVHGLARAGKVGKQTPKVKETKPKPKGRAYKRLLYTRRFVNVLVNGKRRNPGPS 60  
|||||  
RESULT 7  
RS30\_AERPE STANDARD; PRT; 50 AA.  
ID RS30\_AERPE  
AC Q9Y9T9;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 30S ribosomal protein S30.  
GN RPS30E OR APES068.  
OS Aeropyrum pernix.  
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;  
OC Desulfurococaceae; Aeropyrum.

OK NCBI\_TaxID=56636;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=K1;  
 RX MEDLINE=9910339; PubMed=10382966;  
 RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Hatakeyama Y.,  
 RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Anai A., Kosugi H.,  
 RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,  
 RA Takamiya M., Masuda S., Oguchi T., Tanaka T., Kudo Y.,  
 RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kubota K.,  
 RA Nakamura Y., Nomura N., Sake Y., Kikuchi H.,  
 RT "Complete genome sequence of an aerobic hyper-thermophilic  
 RT crenarchaeon, Aeropyrum pernix K1.";  
 RL DNA Ref. 6:83-101(1999).  
 CC -1- SIMILARITY: BELONGS TO THE S10E FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AP00063; BAA8121.1;  
 KM Ribosomal protein; Complete proteome.  
 SQ SEQUENCE 50 AA; 5701 MW; 346A9DABD4E67DCF CRC64;  
 Query Match 30.1%; Score 91.5; DB 1; Length 50;  
 Best Local Similarity 52.3%; Pred. No. 0.00021;  
 Matches 23; Conservative 5; Mismatches 13; Indels 3; Gaps 2;  
 QY 3 HGSILARAGVRCOTPKV-AKOEKKKKTKGRAKRMQYRRFVNV 45  
 DB 4 HGSILKAGVRCOTPKVAKO-KKAYPRRLKRLKQVATIEKV 45  
 RESULT 8  
 ID NCAL\_CWMA5 STANDARD; PRT; 454 AA.  
 AC P03416;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Nucleocapsid protein (Version 1).  
 GN N.  
 OS Murine coronavirus MHV (strain A59).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
 OC Coronaviridae; Coronavirus.  
 OC NCBI\_TaxID=11142;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=83168908; PubMed=6687635;  
 RA Armstrong J., Smeekens S., Spaan W.J.M., Rottier P.J.M.,  
 RA van der Zeijst B.A.M.;  
 RT "Cloning and sequencing of the nucleocapsid and E1 genes of  
 RT coronavirus";  
 RL (in) Rottier P.J.M., van der Zeijst B.A.M., Spaan W.J.M.,  
 RL Horinek M. (eds.);  
 RL Molecular biology and pathogenesis of coronaviruses, pp.155-162,  
 RL Plenum Press, New York (1984).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X00509; CAA25198.1;  
 DR PIR; A04023; VHI9M.  
 DR InterPro; IPR001218; Corona\_nucleocap.  
 DR Pfam; PF00937; Corona\_nucleoca.1.  
 KW Nucleocapsid.  
 SQ SEQUENCE 454 AA; 49728 MW; 4FD75D69BACA822F CRC64;  
 Query Match 22.2%; Score 67.5; DB 1; Length 454;  
 Best Local Similarity 38.2%; Pred. No. 1.2;  
 Matches 21; Conservative 7; Mismatches 24; Indels 3; Gaps 2;  
 QY 6 LARAGVRCOTPKVAKOEKK-KTGRAKRMQYRRFVNVPTFGKKGPAN 58  
 DB 242 LAKLGKDNQGPQVTKQSAKVKRKLTKRQKRTNKKCPVQCGG-KRGRPN 295  
 RESULT 9  
 ID RL4\_MYCPN STANDARD; PRT; 212 AA.  
 AC P75579;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE 50S ribosomal protein L4.  
 GN RPLD OR MP166 OR MP665.  
 OS Mycoplasma pneumoniae.  
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
 OC NCBI\_TaxID=2104;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 29342 / M129;  
 RX MEDLINE=97105885; PubMed=8948633;  
 RA Himmelfeich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,  
 RA Herrmann R.;  
 RT "Complete sequence analysis of the genome of the bacterium Mycoplasma  
 RT pneumoniae";  
 RL Nucleic Acids Res. 24:4420-4449(1996).  
 CC -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY AND SPECIFICALLY TO 23S  
 CC -1- RNA (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE L4P FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE00061; AAB96313.1;  
 DR InterPro; IPR002136; Ribosomal\_L4/L1E.  
 DR Pfam; PF00573; Ribosomal\_L4; 1.  
 KW Ribosomal protein; rRNA-binding; Complete proteome.  
 SQ SEQUENCE 212 AA; 23589 MW; 58421C03A2754D08 CRC64;  
 Query Match 21.7%; Score 66; DB 1; Length 212;  
 Best Local Similarity 37.0%; Pred. No. 0.84;  
 Matches 20; Conservative 7; Mismatches 21; Indels 6; Gaps 2;  
 QY 5 SLARAGVRCOTPKVAKOEKKKKTKGRAKRMQYRRFVNVPTFGKKGPAN 58  
 DB 51 SILTKGEVRGG---GKKPYKQKHTKAKQGSTRNPHFGGIVFQPK--PNNR 98  
 RESULT 10  
 ID NCAL\_CWMA5 STANDARD; PRT; 454 AA.  
 AC P18448;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)

```
DE Nucleocapsid protein (Version 2).
GN N.
OS Murine coronavirus MHV (strain A59).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
OX NCBI_TaxID=11142;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=91021052; PubMed=2171216;
RX Parker M.M., Masters P.S.;
RA "Sequence comparison of the N genes of five strains of the
RT coronavirus mouse hepatitis virus suggests a three domain structure
RT for the nucleocapsid protein.";
RL Virology 179:463-468(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/annouce/
CC or send an email to license@isb-sib.ch).
CC -----
CC ENBL; M35255; AAA46468.1; --
DR PIR; C45340; C45340.
DR InterPro; IPR001218; Corona_nucleocap.
DR Pfam; PF00937; Corona_nucleoca; 1.
KW Nucleocapsid.
SQ SEQUENCE 454 AA; 49587 MW; 4992187AB0EB131E CRC64;
Query Match 21.2%; Score 64.5; DB 1; Length 454;
Best Local Similarity 36.4%; Pred.No.2.7;
Matches 20; Conservative 8; Mismatches 24; Indels 3; Gaps 2;

Qy 6 LARAGKVGQTPTKVAQDEKK--KKKTGAKRMQMQRFRFVNNVPTFGKKKGPNAN 58
|||:|||:|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 242 LAKLGKDGAPQKVTKOSAKEVRQKILNKPKRTPNNKCVPQCFCG-KRGPNQN 295
|||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

RESULT 11
NCAP_CVM3 STANDARD; PRT; 454 AA.
AC P18447;
DT 01-NOV-1990 (Rel. 16, Created)
DD 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-AUG-1991 (Rel. 19, Last annotation update)
DE Nucleocapsid protein.
DE N.
OS Murine coronavirus MHV (strain 3).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
OX NCBI_TaxID=11140;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=91021052; PubMed=2171216;
RX Parker M.M., Masters P.S.;
RA "Sequence comparison of the N genes of five strains of the
RT coronavirus mouse hepatitis virus suggests a three domain structure
RT for the nucleocapsid protein.";
RL Virology 179:463-468(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/annouce/
CC or send an email to license@isb-sib.ch).
CC -----
CC ENBL; M35254; AAA46444.1; --
DR PIR; B45340; B45340.
DR InterPro; IPR001218; Corona_nucleocap.
DR Pfam; PF00937; Corona_nucleoca; 1.
```

KW Nucleocapsid.  
SQ SEQUENCE 454 AA; 49687 MW; 9C46DB2317E3A849 CRC64;

Query Match 21.2%; Score 64.5; DB 1; Length 454;  
Best Local Similarity 36.4%; Pred. No. 2.7;  
Matches 20; Conservative

Qy 6 LARAGKVGTPKVKAKOEKK--KKKTGRAKRMQYNNRRFVNVVPTFGKKKGPNAN 58  
||| ||| ||| ||| ||| :|| :|| :|| :|| :|| :||  
Db 242 LAKLGDAGPKQVTQSAKEVRQKILNKPQRKRTPNKCPCVOOCFG-KRGPPNQ 295

RESULT 12  
NCAP CVMS STANDARD; PRT; 454 AA.  
ID NCAP CVMS  
AC P18449;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE Nucleocapsid protein.  
GN N.  
OS Murine coronavirus MHV (strain S).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
OC Coronaviridae; Coronavirus.  
OX NCBI\_TaxId=11145;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91021052; PubMed=2171216;  
RA Parker M.M., Masters P.S.;  
RL "Sequence comparison of the N genes of five strains of the coronavirus mouse hepatitis virus suggests a three domain structure for the nucleocapsid protein.";  
RC Virology 179:463-468(1990).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - CC the European Bioinformatics Institute. There are no restrictions on its CC use by non-profit institutions as long as its content is in no way CC modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; M35256; AAA46447.1; .  
DR PIR; A45340.  
DR InterPro; IPR001218; Corona\_nucleoacp.  
DR Pfam; PF00937; Corona\_nucleoca; 1.  
KW Nucleocapsid.  
SQ SEQUENCE 454 AA; 49729 MW; 7C4978F20954A227 CRC64;

Query Match 21.2%; Score 64.5; DB 1; Length 454;  
Best Local Similarity 36.4%; Pred. No. 2.7;  
Matches 20; Conservative

Qy 6 LARAGKVGTPKVKAKOEKK--KKKTGRAKRMQYNNRRFVNVVPTFGKKKGPNAN 58  
||| ||| ||| ||| ||| :|| :|| :|| :|| :|| :||  
Db 242 LAKLGDAGPKQVTQSAKEVRQKILNKPQRKRTPNKCPCVOOCFG-KRGPPNQ 295

RESULT 13  
NCAP CVRSD  
ID NCAP CVRSD STANDARD; PRT; 454 AA.  
AC Q02915;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 01-FEB-1994 (Rel. 28, Last annotation update)  
DE Nucleocapsid protein.  
GN N.  
OS Rat coronavirus (strain 681) (SDAV) (Sialodacryoadenitis virus).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
OC Coronaviridae; Coronavirus.  
OX NCBI\_TaxId=33740;  
RN [1]  
RP SEQUENCE FROM N.A.

```

RX MEDLINE=93174973; PubMed=8438589;
RA Kunita S., Mori M., Terada E.;
RT "Sequence analysis of the nucleocapsid protein gene of rat
RT coronavirus SDAV-681."
RL Virology 193:520-523(1993).
CC -1- SIMILARITY: TO NUCLEOCAPSID PROTEINS OF MURINE CORONAVIRUS MHV.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D10760; BAA01591.1; -
DR PIR: A45396; A45396.
DR InterPro: IPR001218; Corona_nucleocap.
DR Pfam: PF00937; Corona_nucleoca; 1.
KW Nucleocapsid.
FT DOMAIN 1 16 SER-RICH.
FT 194 220
SQ SEQUENCE 454 AA; 49437 MW; 247184B05D6CD96D CRC64;

Query Match 21.2%; Score 64.5; DB 1; Length 454;
Best Local Similarity 36.4%; Pred. No. 2.7;
Matches 20; Conservative 8; Mismatches 24; Indels 3; Gaps 2;

Qy 6 LARAGVRCQTPKVAKEKK--KKKTGRAKRMQYNNRFPNVVPTFGKKKGGPNAN 58
Db 242 LAKLGKDAQPQVOTKQSAKEVRQKILNKPQKRTPNKQCPVQCGF-KRGPNN 295

RESULT 14
ID NCAP_CVW1 STANDARD; PRT; 455 AA.
AC P18446;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-AUG-1991 (Rel. 19, Last annotation update)
DE Nucleocapsid protein.
GN N.
OS Murine coronavirus MHV (strain 1).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
OC NCBI_TaxId=11139;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91021052; PubMed=2171216;
RA Parker M.M., Masters P.S.;
RT "Sequence comparison of the N genes of five strains of the
RT coronavirus mouse hepatitis virus suggests a three domain structure
RT for the nucleocapsid protein."
RL Virology 179:463-468(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M35253; AAA46439.1; -
DR PIR: D45340; D45340.
DR InterPro: IPR001218; Corona_nucleocap.
DR Pfam: PF00937; Corona_nucleoca; 1.
KW Nucleocapsid.
SQ SEQUENCE 455 AA; 49694 MW; 382DBE4BD4264BFF CRC64;

Query Match 21.2%; Score 64.5; DB 1; Length 455;
Best Local Similarity 36.4%; Pred. No. 2.7;
Matches 20; Conservative 8; Mismatches 24; Indels 3; Gaps 2;

```

```

Qy 6 LARAGVRCQTPKVAKEKK--KKKTGRAKRMQYNNRFPNVVPTFGKKKGGPNAN 58
Db 242 LAKLGKDAQPQVOTKQSAKEVRQKILNKPQKRTPNKQCPVQCGF-KRGPNN 295

RESULT 15
ID NCAP_CVWJH STANDARD; PRT; 455 AA.
AC P03417;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Nucleocapsid protein.
GN N.
OS Murine coronavirus MHV (strain JHM).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
OC NCBI_TaxId=11144;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8372950; PubMed=6308569;
RA Skinner M.A., Siddell S.G.;
RT "Coronavirus JHM: nucleotide sequence of the mRNA that encodes
RT nucleocapsid protein."
RL Nucleic Acids Res. 11:5045-5054(1983).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X00990; CA425497.1; -
DR PIR: A04024; VHIMJ.
DR InterPro: IPR001218; Corona_nucleocap.
DR Pfam: PF00937; Corona_nucleoca; 1.
KW Nucleocapsid.
SQ SEQUENCE 455 AA; 49714 MW; 36FC529D1272B5BE CRC64;

Query Match 21.2%; Score 64.5; DB 1; Length 455;
Best Local Similarity 36.4%; Pred. No. 2.7;
Matches 20; Conservative 8; Mismatches 24; Indels 3; Gaps 2;

Qy 6 LARAGVRCQTPKVAKEKK--KKKTGRAKRMQYNNRFPNVVPTFGKKKGGPNAN 58
Db 242 LAKLGKDAQPQVOTKQSAKEVRQKILNKPQKRTPNKQCPVQCGF-KRGPNN 295

Search completed: December 3, 2002, 14:32:16
Job time : 12 secs

```





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OW protein - protein search, using SW model

Run on: December 3, 2002, 14:30:48 ; Search time 29 Seconds  
(without alignments)  
419.199 Million cell updates/sec

Title: US-09-424-815E-1  
Perfect score: 304  
Sequence: 1 KVHGLARAGKVRGQTPKVA.....RRFVVVPTFGKKKGPNNAS 59

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL 21: \*  
1: sp\_archaea: \*  
2: sp\_bacteria: \*  
3: sp\_fungi: \*  
4: sp\_human: \*  
5: sp\_invertebrate: \*  
6: sp\_mammal: \*  
7: sp\_mhc: \*  
8: sp\_organelle: \*  
9: sp\_phage: \*  
10: sp\_plant: \*  
11: sp\_rodent: \*  
12: sp\_virus: \*  
13: sp\_vertebrate: \*  
14: sp\_unclassified: \*  
15: sp\_virus: \*  
16: sp\_bacteria: \*  
17: sp\_archaea: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	304	100.0	133	4	Q9H5V4	Q9H5V4 homo sapien
2	304	100.0	133	11	Q9J2Z4	Q9J124 mus musculus
3	304	100.0	137	11	Q920W8	Q920W8 mus musculus
4	304	100.0	137	11	Q920W7	Q920W7 mus spicile
5	304	100.0	137	11	Q91V99	Q91V99 mus musculus
6	288	94.7	133	13	Q90YPI	Q90YPI ictaluruss p
7	269	88.5	131	5	Q962Q1	Q962Q1 spodioplera
8	263	86.5	132	5	Q9VDM8	Q9VDM8 drosofila
9	231	76.0	92	10	Q9M0E4	Q9M0E4 arabidopsis
10	227	74.7	130	5	Q18231	Q18231 caenorhabdi
11	112	36.8	230	4	Q15351	Q15351 homo sapien
12	104.5	34.4	52	17	Q972H4	Q972H4 sulfolobus
13	101.5	33.4	55	17	Q971W4	Q971W4 sulfolobus
14	95	31.2	59	5	Q8SR05	Q8SR05 encephalito
15	87.5	28.8	55	17	Q8ZV51	Q8ZV51 pyrobaculum
16	65	21.4	219	10	Q9LGS6	Q9LGS6 Oryza sativi

17	64.5	21.2	425	12	Q83357	Q83357 murine hepa
18	64.5	21.2	451	12	Q83359	Q83359 murine hepa
19	64.5	21.2	454	12	Q83358	Q83358 murine hepa
20	64.5	21.2	454	12	Q9WCD0	Q9WCD0 rat coronav
21	64.5	21.2	454	12	Q9J3E2	Q9J3E2 murine hepa
22	64.5	21.2	455	12	Q83360	Q83360 murine hepa
23	64.5	21.2	455	12	Q9QC29	Q9QC29 murine hepa
24	64.5	21.2	457	12	Q08614	Q08614 murine hepa
25	64	21.1	425	16	Q8YX55	Q8YX55 anabena sp
26	63.5	20.9	451	12	Q9J3P6	Q9J3P6 murine hepa
27	63.5	20.9	451	12	Q72588	Q72588 murine hepa
28	63.5	20.9	451	12	Q9PY96	Q9PY96 murine hepa
29	63.5	20.9	1510	5	Q25920	Q25920 plasmidium
30	63	20.7	260	10	Q9LQZ9	Q9LQZ9 arabidopsis
31	63	20.7	290	4	Q99463	Q99463 homo sapien
32	63	20.7	421	5	Q9BL55	Q9BL55 caenorhabdi
33	63	20.7	517	10	Q9LKR2	Q9LKR2 arabidopsis
34	62.5	20.6	454	12	Q9IKC6	Q9IKC6 rat saloda
35	62	20.4	467	16	Q8R617	Q8R617 t hypotheti
36	62	20.4	917	5	Q9VBX5	Q9VBX5 drosophila
37	61.5	20.2	171	10	Q9LDA1	Q9LDA1 oryza sativ
38	61.5	20.2	836	11	Q9WY8	Q9WY8 rattus norv
39	61	20.1	317	5	Q9U613	Q9U613 atemia san
40	60.5	19.9	351	11	Q9CW80	Q9CW80 mus musculu
41	60.5	19.9	731	4	Q60433	Q60433 homo sapien
42	60	19.7	115	11	Q64390	Q64390 rattus norv
43	60	19.7	435	10	Q940Y0	Q940Y0 arabidopsis
44	60	19.7	521	3	Q42650	Q42650 echinosach
45	60	19.7	893	3	Q8WZ55	Q8WZ55 neurospora

#### ALIGNMENTS

##### RESULT 1

Q9H5V4 PRELIMINARY; PRT; 133 AA.  
AC Q9H5V4;  
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 19, last annotation update)  
DE CDNA: FLJ22986 fis, clone KAT11742.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,  
RA Suzuki Y., Ohtsuka M., Nishii T., Shibahara T., Tanaka T.,  
RA Nakamura Y., Iwagaki T., Sugano S.,  
RT "NEO human cDNA sequencing project."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK026639; BAB15515.1; -  
DR HSSP; P02248; IUBI.  
DR InterPro; IPR00626; Ubiquitin.  
DR Pfam; PF00240; ubiquitin.1.  
DR PRINTS; PR00348; Ubiquitin.  
DR SMART; SM00213; UBO; 1.  
DR PROSITE; PS00299; Ubiquitin\_1; 1.  
DR PROSITE; PS0053; Ubiquitin\_2; 1.  
SQ SEQUENCE 133 AA; 14390 MW; 5D2F81F2A3558559 CRC64;

Query Match 100.0%; Score 304; DB 4; Length 133;  
Best Local Similarity 100.0%; Pred. No. 4.8e-30;  
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVHGLARAGKVRGQTPKVAKKKKKTKGRKRRMVRFRVNVPTFGKKKGPNNAS 59  
DB 75 KVHGLARAGKVRGQTPKVAKKKKKTKGRKRRMVRFRVNVPTFGKKKGPNNAS 133  
RESULT 2



DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Fau protein (Fragment).  
 GN FAU.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VARIOUS STRAINS;  
 RA Liu Y., Kitano T., Koide T., Shirosaki T., Moriaki K., Saitou N.;  
 RT "Consipuous Differences among Gene Genealogies of 21 Nuclear Genes of  
 RT Five Mus musculus subspecies."  
 RL Submitted (FE8-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB039084; BAB6608.1; -;  
 DR EMBL; AB039085; BAB6609.1; -;  
 DR EMBL; AB039087; BAB6611.1; -;  
 DR EMBL; AB039088; BAB6612.1; -;  
 DR EMBL; AB039089; BAB6613.1; -;  
 DR EMBL; AB039090; BAB6614.1; -;  
 DR EMBL; AB039091; BAB6615.1; -;  
 DR EMBL; AB039092; BAB6616.1; -;  
 DR InterPro; IPR000626; Ubiquitin.  
 DR Pfam; PF00240; Ubiquitin.1.  
 DR PROSITE; PS00299; UBIQUITIN\_1; UNKNOWN\_1.  
 DR PROSITE; PS0053; UBIQUITIN\_2; 1.  
 FT NON\_TER  
 SQ SEQUENCE 137 AA; 14787 MW; 57099FF7065D8828 CRC64;

Query Match 100.0%; Score 304; DB 11; Length 137;  
 Best Local Similarity 100.0%; Pred. No. 5e-30; Indels 0; Gaps 0;  
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KHGSLARAGKVGQTPKVAKEKKKKTKGRARRMQYNRRFVNVVPTFGKKKGNPNS 59  
 Db 79 KHGSLARAGKVGQTPKVAKEKKKKTKGRARRMQYNRRFVNVVPTFGKKKGNPNS 137

RESULT 6  
 Q90YPI PRELIMINARY; PRT; 133 AA.  
 AC Q90YPI;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DE 40S ribosomal protein S30.  
 OS Ictaelurus punctatus (Channel catfish).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes;  
 OC Ictaluridae; Ictalurus.  
 OX NCBI\_TaxId=7998;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Karel A., Patterson A., Feng J., Liu Z.J.;  
 RT "translational machinery of channel catfish: I. A transcriptomic  
 RT approach to the analysis of 32 40S ribosomal protein genes and their  
 RT expression."  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF402841; AAK95215.1; -;  
 DR InterPro; IPR000626; Ubiquitin.  
 DR Pfam; PF00240; Ubiquitin.1.  
 DR PROSITE; PS0053; UBIQUITIN\_2; 1.  
 KW Ribosomal protein.  
 SQ SEQUENCE 133 AA; 14504 MW; 62036B0E72C5CAC CRC64;

Query Match 94.7%; Score 288; DB 13; Length 133;  
 Best Local Similarity 94.9%; Pred. No. 4.6e-28; Indels 0; Gaps 0;  
 Matches 56; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 KHGSLARAGKVGQTPKVAKEKKKKTKGRARRMQYNRRFVNVVPTFGKKKGNPNS 59

Db 75 KHGSLARAGKVGQTPKVAKEKKKKTKGRARRMQYNRRFVNVVPTFGKKKGNPNS 133  
 RESULT 7  
 Q96Z01 PRELIMINARY; PRT; 131 AA.  
 AC Q96Z01;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Ribosomal protein S30 (Fall armyworm).  
 OS Spodoptera frugiperda (Fall armyworm).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;  
 OC Noctuidae; Noctuidae; Amphipyridae; Spodoptera.  
 OX NCBI\_TaxId=7108;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Landais I., Ogliastro M., Mita K., Nohata J., Lopez-Farber M.,  
 RA Duonot-Cerutti M., Fournier P., Devauchelle G.;  
 RT "Full-length ribosomal protein sequence from an EST library of  
 RT Spodoptera frugiperda cells (Sf9)."  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF400225; AAK92197.1; -;  
 SQ SEQUENCE 131 AA; 14314 MW; 03AEAD0E31EBC1B04 CRC64;

Query Match 88.5%; Score 269; DB 5; Length 131;  
 Best Local Similarity 84.7%; Pred. No. 1e-25; Indels 0; Gaps 0;  
 Matches 50; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KHGSLARAGKVGQTPKVAKEKKKKTKGRARRMQYNRRFVNVVPTFGKKKGNPNS 59  
 Db 73 KHGSLARAGKVGQTPKVAKEKKKKTKGRARRMQYNRRFVNVVPTFGRRGPNNS 131

RESULT 8  
 Q9VDH8 PRELIMINARY; PRT; 132 AA.  
 AC Q9VDH8;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE CG15697 protein (RH08962p).  
 GN CG15697.  
 OS Drosophila melanogaster (Fruit Fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxId=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; Pubmed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutten G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miles G.L.G.,  
 RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,  
 RA Ballaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Buck J., Brockstein P., Brothier P.,  
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fessler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,



RT *elegans*; (2)  
RL Nature 368:32-38(1994).  
RN (2)  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Geisel C., Stellyes L., Bradshaw H.;  
RT "The sequence of *C. elegans* cosmid C26F1.";  
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
RN (3)  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Waterston R.;  
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U53148; AAB37076.1; -.  
DR HSSP; P02248; IUBI.  
DR InterPro; IPR000626; Ubiquitin.  
DR Pfam; PF00240; ubiquitin; 1.  
DR SMART; SM00213; UBO; 1.  
SQ SEQUENCE 130 AA; 14033 MW; 50DC09AFB9P48532 CRC64;

Query Match 74.7%; Score 227; DB 5; Length 130;  
Best Local Similarity 76.3%; Pred. No. 1.6e-20;  
Matches 45; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy 1 KHGSLARAGKVRGOTPKVAKOEKKKKTGRAKRMQYNRRFVNVVPTFGKKGGPNNS 59  
Db 72 KHGSLARAGKVAQTPKVDKDKKKKGRAPFRVGYTRRYNVVAGSGPKKGGPNNS 130

RESULT 11  
ID 015351 PRELIMINARY; PRT; 230 AA.  
AC 015351;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE SEB4B (Fragment).  
GN SEB4B (HUMAN).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OC NCBI\_TaxID=9606;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC TISSUE=THYMUS;  
RA Ruehlmann A., Gupta A., Terhorst C.;  
RT "A novel murine RRM-type protein and its human homolog.";  
RL Submitted (SEP-1993) to the EMBL/GenBank/DBJ databases.  
DR EMBL; X75315; CA53064.1; -.  
DR HSSP; P09651; IUP1.  
DR InterPro; IPR000504; RNA\_rec\_mot.  
DR Pfam; PF00076; Rrm; 1.  
DR SMART; SM00360; RRM; 1.  
DR PROSITE; PS50102; RRM; 1.  
DR PROSITE; PS00030; RRM\_RNP\_1; UNKNOWN\_1.  
FT NON TER 1  
SQ SEQUENCE 230 AA; 25220 MW; C747D6500608461 CRC64;

Query Match 36.8%; Score 112; DB 4; Length 230;  
Best Local Similarity 100.0%; Pred. No. 5.3e-06;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 34 RRMQYNRRFVNVVPTFGKKG 54  
Db 1 RRMQYNRRFVNVVPTFGKKG 21

RESULT 12  
Q972H4  
ID 0972H4 PRELIMINARY; PRT; 52 AA.  
AC 0972H4;  
DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)

DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE LSU ribosomal protein S30E (rps30E).  
GN RPS30E OR SSO6817.  
OS Sulfolobus solfataricus.  
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
OC Sulfolobus.  
RN (1)  
NCBI\_TaxID=2287;  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 35092 / DSM 1617 / P2;  
RX MEDLINE=21332296; PubMed=11427726;  
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,  
RA Aways M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,  
RA De Moers A., Brauso G., Fletcher C., Gordon P.M.K.,  
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,  
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,  
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,  
RA Garrett R.A., Ragan M.A., Senses C.W., Van der Oost J.;  
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).  
DR EMBL; AE006714; AAK41214.1; -.  
KW Ribosomal protein; Complete proteome.  
SQ SEQUENCE 52 AA; 6046 MW; A9EB9D3E4E7E744E CRC64;

Query Match 34.4%; Score 104.5; DB 17; Length 52;  
Best Local Similarity 51.2%; Pred. No. 9.2e-06;  
Matches 21; Conservative 7; Mismatches 12; Indels 1; Gaps 1;

Qy 3 HGSILARAGKVRGOTPKVAKOEKKKKTGRAKRMQYNRRV 43  
Db 4 HGSILTRAGKVRGOTPKIQPK-KHKEVPRVNRKYEKRV 43

RESULT 13  
ID 0971W4 PRELIMINARY; PRT; 55 AA.  
AC 0971W4;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Putative 30S ribosomal protein S30.  
GN STS139.  
OS Sulfolobus tokodaii.  
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
OC Sulfolobus.  
RN (1)  
NCBI\_TaxID=111955;  
RP SEQUENCE FROM N.A.  
RC STRAIN=JCM 10545 / 7;  
RX PubMed=11572479;  
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,  
RA Sekine M., Baba S.-I., Anka H., Kosugi H., Hasegawa A., Fukui S.,  
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,  
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A.,  
RA Aoki K.-I., Maeuda S., Yanagii M., Nishimura M., Yamagishi A.,  
RA Oshima T., Kikuchi H.;  
RT "Complete genome sequence of an aerobic thermophilic  
RT Crenarchaeon, Sulfolobus tokodaii strain7.";  
RL DNA Res. 8:123-140(2001).  
DR EMBL; AP000985; BAB66306.1; -.  
KW Ribosomal protein; Hypothetical protein; Complete proteome.  
SQ SEQUENCE 55 AA; 6365 MW; 5C9404787BA65826 CRC64;

Query Match 33.4%; Score 101.5; DB 17; Length 55;  
Best Local Similarity 48.8%; Pred. No. 2.3e-05;  
Matches 20; Conservative 9; Mismatches 11; Indels 1; Gaps 1;

Qy 3 HGSILARAGKVRGOTPKVAKOEKKKKTGRAKRMQYNRRV 43  
Db 4 HGSILTRAGKVRGOTPKPKKE-RHKEVPRVNRKYEKRV 43

RESULT 14

Search completed: December 3, 2002, 14:32:53  
Job time : 32 secs

```
Q8SR05
ID Q8SR05      PRELIMINARY;      PRT;      59 AA.
AC Q8SR05;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE 40S ribosomal protein S30.
GN ECU10_1575.
OS Encephalitozoon cuniculi.
OC Eukaryota; Microsporidia; Unikaryonidae; Encephalitozoon.
OX NCBI_TaxID=6035;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GB-M1;
RA Genoscope;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=GB-M1;
RX MEDLINE=21576510; PubMed=11719806;
RA Katinka M.D., Duprat S., Cornillot E., Metenier G., Thomarat F.,
RA Prensier G., Barbe V., Peyretailade E., Brottier P., Wincker P.,
RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,
RA Weissenbach J., Vivares C.P.;
RA "Genome sequence and gene compaction of the eukaryote parasite
RT Encephalitozoon cuniculi.";
RL Nature 414:450-453(2001).
DR EMBL; AL590449; CAD25878.1; -.
SQ SEQUENCE 59 AA; 6815 MW; C3A4A9AB33488FDB CRC64;

Query Match      31.2%; Score 95; DB 5; Length 59;
Best Local Similarity 45.9%; Pred. No. 0.00016;
Matches 17; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 5 SLARAGKVRGQTPKVAQOEKKKKTKTGRAKRRMQYNRR 41
Db 5 TINKAGKVRNQTDPKDPVVEKERKKCGRCRQLKPEKR 41

RESULT 15
Q8ZV51
ID Q8ZV51      PRELIMINARY;      PRT;      55 AA.
AC Q8ZV51;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Ribosomal protein S30.
GN PAE2455.
OS Pyrobaculum aerophilum.
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC Thermoproteaceae; Pyrobaculum.
OX NCBI_TaxID=13773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IM2 / ATCC 51768 / DSM 7523;
RX PubMed=11792869;
RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA Miller J.H.;
RA "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
RT aerophilum.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
DR EMBL; AE009877; AAL64205.1; -.
KW Complete proteome.
SQ SEQUENCE 55 AA; 6200 MW; E2F91102FE9EFF90 CRC64;

Query Match      28.8%; Score 87.5; DB 17; Length 55;
Best Local Similarity 48.7%; Pred. No. 0.0012;
Matches 19; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

QY 3 HGSLLARAGKVRGQTPKVAQOEKKKKTKTGRAKRRMQYNRR 41
Db 4 HGSLLTAKAGKVRNQTPKI--PAKPRKNLTTPRRNIRNYKER 41
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Comphen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2002, 14:30:49 / Search time 14 Seconds  
(without alignments)  
123.997 Million cell updates/sec

Title: US-09-424-815E-1  
Perfect score: 304  
Sequence: 1 KXHSLARAGKVGQTPKVA.....RRFVNVPFCKKGNPANS 59

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/1aa/backfill1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	63	20.7	370	1	US-08-415-818-12 Sequence 12, Appl
2	63	20.7	370	2	US-08-894-236-12 Sequence 12, Appl
3	63	20.7	370	5	PCT-US96-01444-12 Sequence 12, Appl
4	59.5	19.6	315	4	US-09-615-192A-378 Sequence 378, App
5	54.5	17.9	171	4	US-09-134-001C-5286 Sequence 5286, Ap
6	54	17.8	290	2	US-08-903-801-1 Sequence 1, Appl1
7	54	17.8	290	4	US-09-295-055-1 Sequence 1, Appl1
8	53.5	17.6	747	4	US-09-724-864-36 Sequence 36, Appl
9	53.5	17.6	714	2	US-08-990-114-3 Sequence 3, Appl1
10	53.5	17.6	714	4	US-09-241-333-3 Sequence 3, Appl1
11	53	17.4	318	4	US-09-615-192A-389 Sequence 389, App
12	53	17.4	1253	4	US-07-920-281C-3 Sequence 3, Appl1
13	53	17.4	1253	4	US-08-466-277-3 Sequence 3, Appl1
14	52.5	17.3	135	4	US-09-134-001C-3305 Sequence 11, Appl
15	52.5	17.3	351	4	US-09-245-041-11 Sequence 57, Appl
16	52.5	17.3	427	1	US-08-476-008-57 Sequence 57, Appl
17	52.5	17.3	427	1	US-08-306-063-57 Sequence 57, Appl
18	52.5	17.3	427	1	US-08-833-485-57 Sequence 57, Appl
19	52.5	17.3	427	4	US-09-243-374-8 Sequence 8, Appl1
20	52.5	17.3	427	4	US-09-137-440-57 Sequence 57, Appl
21	52.5	17.3	427	6	5310667-8 Patent No. 5310667
22	52	17.1	25	4	US-09-039-780A-119 Sequence 119, App
23	52	17.1	28	1	US-08-384-212-1 Sequence 1, Appl1
24	52	17.1	28	2	US-08-739-819-1 Sequence 1, Appl1
25	52	17.1	35	1	US-08-384-212-2 Sequence 2, Appl1
26	52	17.1	35	2	US-08-739-819-2 Sequence 2, Appl1
27	52	17.1	88	4	US-09-134-001C-2855 Sequence 2855, Ap

28	52	17.1	139	3	US-08-464-841A-2 Sequence 2, Appl1
29	52	17.1	143	3	US-08-464-841A-4 Sequence 4, Appl1
30	52	17.1	154	4	US-08-871-732A-1 Sequence 1, Appl1
31	52	17.1	154	4	US-09-346-510B-1 Sequence 1, Appl1
32	51.5	16.9	59	2	US-08-460-890A-60 Sequence 60, Appl
33	51.5	16.9	59	3	US-08-167-641C-60 Sequence 60, Appl
34	51.5	16.9	59	4	US-08-460-971A-60 Sequence 60, Appl
35	51.5	16.9	59	4	US-08-462-040-60 Sequence 60, Appl
36	51.5	16.9	472	2	US-08-216-894-10 Sequence 10, Appl
37	51.5	16.9	472	4	US-09-115-746-10 Sequence 10, Appl
38	51.5	16.9	801	4	US-09-104-070-2 Sequence 2, Appl1
39	51.5	16.9	1202	1	US-08-425-061-22 Sequence 22, Appl
40	51.5	16.9	1202	2	US-08-825-886-22 Sequence 22, Appl
41	51.5	16.9	1363	1	US-08-425-061-23 Sequence 23, Appl
42	51.5	16.9	1363	2	US-08-825-886-23 Sequence 23, Appl
43	51.5	16.9	1852	1	US-08-425-061-24 Sequence 24, Appl
44	51.5	16.9	1852	2	US-08-825-886-24 Sequence 24, Appl
45	51.5	16.9	1863	1	US-08-425-061-16 Sequence 16, Appl

ALIGNMENTS

RESULT 1  
US-08-415-818-12  
Sequence 12, Application US/08415818  
Patent No. 5621079  
GENERAL INFORMATION:  
APPLICANT: Cascieri, Margaret A.  
APPLICANT: Linemeyer, David L.  
APPLICANT: MacNeill, Douglas J.  
APPLICANT: Shiao, Lin-Lin  
APPLICANT: Strader, Catherine D.  
APPLICANT: Tan, Carina P.  
APPLICANT: Weinberg, David H.  
TITLE OF INVENTION: NEUROPEPTIDE Y RECEPTOR  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Mary A. Appollina  
STREET: P.O. Box 2000, 126 E. Lincoln Ave.  
CITY: Rahway  
STATE: NJ  
COUNTRY: USA  
ZIP: 07065  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/415,818  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/383,746  
FILING DATE: 03-FEB-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Appollina, Mary A.  
REGISTRATION NUMBER: 34,087  
REFERENCE/DOCKET NUMBER: 19390  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 908-594-3462  
TELEFAX: 908-594-4720  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 370 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-415-818-12  
Query Match 20.7%; Score 63; DB 1; Length 370;





;; PRIOR FILING DATE: 1998-10-09  
;; NUMBER OF SEQ ID NOS: 405  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 378  
;; LENGTH: 315  
;; TYPE: PRT  
;; ORGANISM: Eucalyptus grandis  
US-09-615-192A-378

Query Match 19.6%; Score 59.5; DB 4; Length 315;  
Best Local Similarity 31.8%; Pred. No. 4.1;  
Matches 21; Conservative 8; Mismatches 18; Indels 19; Gaps 3;

QY 5 SLARAGKVRGQTPKVAKOEKKKKTKGRARMOYRRFVN-----VPTF--GKKK 53  
DB 37 SLVRAG-----VAKAIENETRTGASLLRLHPHDCFVAGCDASILLDPTSPFVGKTA 88

QY 54 GPNANS 59  
DB 89 APNNNS 94

RESULT 5  
US-09-134-001C-5286  
; Sequence 5286, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 5286  
; LENGTH: 171  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-5286

Query Match 17.9%; Score 54.5; DB 4; Length 171;  
Best Local Similarity 25.8%; Pred. No. 9.4;  
Matches 17; Conservative 12; Mismatches 20; Indels 17; Gaps 2;

QY 6 LARAGK-----VRGQTPKVAKOEKKKKTKGRAR--RMOYRRFVNVPPT 48  
DB 106 LVGAGSLKRLQKQLERARYEVGSDYERMKRKLAKQKRREAOXKOTRRFMSLRN 165

QY 49 FGKKGK 54  
DB 166 TSORCG 171

RESULT 6  
US-08-903-801-1  
; Sequence 1, Application US/08903801  
; Patent No. 5932712  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Shah, Puri  
; TITLE OF INVENTION: NEW ANEXIN BINDING PROTEIN  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA

;; ZIP: 94304  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette  
;; COMPUTER: IBM Compatible  
;; OPERATING SYSTEM: DOS  
;; SOFTWARE: FastSeq for Windows Version 2.0  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/903,801  
;; FILING DATE: Herewith  
;; CLASSIFICATION: 536  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER:  
;; FILING DATE:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Billings, Lucy J.  
;; REGISTRATION NUMBER: 36,749  
;; REFERENCE/DOCKET NUMBER: PF-0354 US  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 415-855-0555  
;; TELEFAX: 415-845-4166  
;; TELEX:  
;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 290 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; IMMEDIATE SOURCE:  
;; LIBRARY: PROSONO1  
;; CLONE: 2272281  
US-08-903-801-1

Query Match 17.8%; Score 54; DB 2; Length 290;  
Best Local Similarity 26.3%; Pred. No. 19;  
Matches 15; Conservative 10; Mismatches 10; Indels 22; Gaps 2;

QY 9 AGKVRGQTPKVA-----KOEKKKKTKGRARMOYRRFVNVPPTFGKKKGN 56  
DB 208 ASEEKAEPTPAEDNDGCDKKKKKKKKGEKEKEKE-----KKKGS 251

RESULT 7  
US-09-295-055-1  
; Sequence 1, Application US/09295055  
; Patent No. 6232440  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Shah, Puri  
; TITLE OF INVENTION: NEW ANEXIN BINDING PROTEIN  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/295,055  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/903,801  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749



```
APPLICATION NUMBER: 08/990,114
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0451 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 714 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 128842
US-09-241-333-3
```

Query Match 17.6%; Score 53.5; DB 4; Length 714;  
Best Local Similarity 23.7%; Pred. No. 60;  
Matches 18; Conservative 9; Mismatches 20; Indels 29; Gaps 1;

```
QY 6 LARAGVVGQTPKVA-----KOEKKKKTGRAKRRM 36
DB 4 LAKAGTKEAKMAHPKVEDESEDESEBDDSSGSEVVIPOKKGKATATPAKKV 63
QY 37 QYNRRFVNVPFGKK 52
DB 64 VVSQTKKVAVPTPAKK 79
```

```
RESULT 11
US-09-615-192A-389
Sequence 389, Application US/09615192A
Patent No. 6410718
GENERAL INFORMATION:
APPLICANT: Bloksberg, Leonard N.
TITLE OF INVENTION: Materials and Methods for the
FILE REFERENCE: 11000.1003c4U
CURRENT APPLICATION NUMBER: US/09/615,192A
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 08/975,316
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: US 08/713,000
PRIOR FILING DATE: 1996-09-11
PRIOR APPLICATION NUMBER: US 09/169,789
PRIOR FILING DATE: 1998-10-09
NUMBER OF SEQ ID NOS: 405
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 389
LENGTH: 318
TYPE: PRT
ORGANISM: Pinus radiata
US-09-615-192A-389
```

Query Match 17.4%; Score 53; DB 4; Length 318;  
Best Local Similarity 27.6%; Pred. No. 29;  
Matches 21; Conservative 10; Mismatches 27; Indels 18; Gaps 3;

```
QY 2 VHGSLARA--GKVRGQTPKVAQ-----EKKKKKTGRAKRRMOYRRFV----- 43
DB 22 VNGQLSTFYAKSCPRIPISVVKQAVAKKMGASLVRLHFHDFCVNGCDSILLDD 81
QY 44 NVVPTFGKKKGNANS 59
DB 82 NATFTGKTAGPNANS 97
```

```
RESULT 12
US-07-920-281C-3
Sequence 3, Application US/07920281C
Patent No. 5739026
GENERAL INFORMATION:
APPLICANT: Garoff, Henrik
TITLE OF INVENTION: DNA Expression Systems Based on
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/920,281C
FILING DATE: 13-AUG-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 828-103P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-241-1300
TELEFAX: 703-241-2848
TELEX: 248345
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1253 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-920-281C-3
```

Query Match 17.4%; Score 53; DB 1; Length 1253;  
Best Local Similarity 44.0%; Pred. No. 1.3e+02;  
Matches 11; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

```
QY 10 GKVRGQTPKVAKOEKKKKTGRAKR 34
DB 86 GKTQQKKKKQADKKKKKPKRRR 110
```

```
RESULT 13
US-08-466-277-3
Sequence 3, Application US/08466277
Patent No. 6190666
GENERAL INFORMATION:
APPLICANT: Garoff, Henrik
TITLE OF INVENTION: DNA Expression Systems Based on
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,277
FILING DATE: 06-Jun-1995
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/920,281
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 828-103P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-241-1300
TELEFAX: 703-241-2848
TELEX: 248345
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1253 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-08-466-277-3

```

Query Match 17.4%; Score 53; DB 4; Length 1253;  
Best Local Similarity 44.0%; Pred. No. 1.3e+02;  
Matches 11: Conservative 5; Mismatches 9; Indels

Qy 10 GKVRGQTPKVAQEKKKKKTGRAKR 34  
||:| |:| |:| |:| |:|  
Db 86 GKTQQQKKKDKQADKKKKKPKGRER 110

RESULT 14

US-09-134-001C-3305  
; Sequence 3305, Application US/09134001C  
; Patent No. 6380370

**GENERAL INFORMATION:**

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC-007

FILE REFERENCE: GIC-007  
CURRENT APPLICATION NUMBER: US/09/134, 001C

; CURRENT AFFILIATION NUMBER: 03/0  
 ; CURRENT FILING DATE: 1998-08-13

; CURRENT FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08

; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779

PRIOR FILING DATE: 1997-08-14

;; PRIOR FILING DATE: 1997-08  
;; NUMBER OF SEQ ID NOS: 5674

; NUMBER OF SEQ ID  
: SEQ ID NO 3305

```

; SEQ ID NO 330
; LENGTH: 135

```

```

; LENGTH: 1-
; TYPE: PRT

```

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-3305

Query Match 17.3%; Score 52.5; DB 4; Length 135;  
Best Local Similarity 36.8%; Pred. No. 13;  
Matches 14: Conservative 11; Mismatches 10; Indels

**Qy**            4 GSLRAGKVRGQTPKVAKQEKKKKTKRAKRMMQYNRR 41  
             ||||| : : : : : : : : : : : : : : : :  
**Db**            101 GSLKRGALTT-RDPRM--KEPKPGLKARRSPFSKR 135

RESULT 15

US-09-245-041-11

03-03-243-041-11  
: Sequence 11, Application US/09245041

; Sequence 11, Appendix 11; Patent No. 6274339

FILE NO. 0274333  
; GENERAL INFORMATION:  
; GENERAL INFORMATION:

APPLICANT: MOORE, K.

APPLICANT: Nagle, D.

APPLICANT: NAME, D.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND TREATMENT  
; TITLE OF INVENTION: OF BODY WEIGHT DISORDERS INCLUDING OBESITY

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Comphen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2002, 14:31:22 ; Search time 11 Seconds  
(without alignments)  
85.412 Million cell updates/sec

Title: US-09-424-815E-1

Perfect score: 304  
Sequence: 1 KVHSLARAGKVRGQTPKVA.....RRFNVVPTFGKKKGNANS 59

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 102317 seqs, 15924203 residues

Total number of hits satisfying chosen parameters: 102317

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	304	100.0	170	10	US-09-925-301-994
2	276	90.8	118	10	US-09-864-761-35333
3	62.5	20.6	206	10	US-09-214-881A-7
4	61.4	20.1	185	10	US-09-214-881A-9
5	60	19.7	86	10	US-09-864-761-31832
6	59	19.4	66	10	US-09-764-869-722
7	59	19.4	66	10	US-09-764-846-163
8	58	19.1	72	10	US-09-864-761-36199
9	58	19.1	208	10	US-09-214-881A-2
10	58	19.1	209	10	US-09-214-881A-6
11	58	19.1	209	10	US-09-214-881A-8
12	57	18.8	79	10	US-09-764-846-193
13	57	18.8	170	10	US-09-864-761-39624
14	56.5	18.6	54	10	US-09-864-761-45531
15	55.5	18.3	98	10	US-09-764-846-229
16	55.5	18.3	241	10	US-09-938-803-8
17	55.5	18.3	143	10	US-09-946-805-4
18	55	18.1	533	10	US-09-431-226-5
19	55	18.1	747	10	US-09-431-226-7

20	54.5	17.9	63	10	US-09-764-846-185	Sequence 185, App
21	54.5	17.9	72	10	US-09-864-761-35548	Sequence 35548, A
22	54.5	17.9	76	10	US-09-764-869-960	Sequence 960, App
23	54.5	17.9	76	10	US-09-764-846-257	Sequence 257, App
24	54.5	17.9	86	10	US-09-864-761-46660	Sequence 46660, A
25	54.5	17.9	150	10	US-09-764-846-225	Sequence 225, App
26	54	17.8	71	10	US-09-864-761-37061	Sequence 37061, A
27	54	17.8	76	9	US-10-002-344A-229	Sequence 229, App
28	54	17.8	141	10	US-09-864-761-36181	Sequence 36181, A
29	54	17.8	290	10	US-09-808-885-1	Sequence 1, Appli
30	54	17.8	525	10	US-09-814-550-2	Sequence 2, Appli
31	53.5	17.6	58	10	US-09-764-846-261	Sequence 261, App
32	53.5	17.6	61	10	US-09-764-846-189	Sequence 189, App
33	53.5	17.6	83	10	US-09-864-761-34889	Sequence 34889, A
34	53.5	17.6	130	10	US-09-815-242-12016	Sequence 12016, A
35	53.5	17.6	130	10	US-09-815-242-13732	Sequence 13732, A
36	53.5	17.6	369	10	US-09-925-300-1070	Sequence 1070, Ap
37	53.5	17.6	498	10	US-09-764-864-1122	Sequence 1122, Ap
38	53.5	17.6	714	10	US-09-978-242-3	Sequence 3, Appli
39	53	17.4	57	10	US-09-864-761-41286	Sequence 41286, A
40	53	17.4	63	10	US-09-764-846-260	Sequence 260, App
41	53	17.4	1253	10	US-09-901-106-3	Sequence 3, Appli
42	52.5	17.3	45	10	US-09-764-846-258	Sequence 258, App
43	52.5	17.3	351	10	US-09-893-238-11	Sequence 11, Appli
44	52.5	17.3	427	9	US-09-464-099A-57	Sequence 57, Appli
45	52.5	17.3	427	10	US-09-861-696-57	Sequence 57, Appli

## ALIGNMENTS

RESULT 1  
US-09-925-301-994  
; Sequence 994, Application US/09925301  
; Patent No. US20020052308A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA106  
; CURRENT APPLICATION NUMBER: US/09/925,301  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05882  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; NUMBER OF SEQ ID NOS: 1694  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 994  
; LENGTH: 170  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (5)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-301-994

Query Match 100.0%; Score 304; DB 10; Length 170;  
Best Local Similarity 100.0%; Pred. No. 3.6e-29;  
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVHSLARAGKVRGQTPKVAKKKKKTKGRARRRQVRRFNVVPTFGKKKGNANS 59  
DB 112 KVHSLARAGKVRGQTPKVAKKKKKTKGRARRRQVRRFNVVPTFGKKKGNANS 170

RESULT 2  
US-09-864-761-35333  
; Sequence 35333, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.



```

OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 8.3
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 6.8
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 13
OTHER INFORMATION: EXPRESSED IN B1474, SIGNAL = 6.5
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 12
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 6.3
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 9.1
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.1
US-09-864-761-33632

Query Match      19.7% Score 60; DB 10; Length 86;
Best Local Similarity 38.2% Pred. No. 1;
Matches 13; Conservative 8; Mismatches 13; Indels 0; Gaps 0.

QY      8 RAGKVGQTPKVAQKQKKKTGRARRMOYNNR 41
      |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      12 RGRRRRRKKKKKKKKKKKKKKKKRRRRRRRR 45

RESULT 6
US-09-764-869-722
; Sequence 722, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 722
; LENGTH: 66
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (4)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (56)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-869-722

Query Match      19.4% Score 59; DB 10; Length 66;
Best Local Similarity 34.9% Pred. No. 1;
Matches 15; Conservative 8; Mismatches 12; Indels 8; Gaps 1.

QY      1 KVHGLARAG-----KVGQTPKVAQKQKKKTGRARR 35
      |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      20 RVRSRSRAAEMTFSSKKKKKKKKKKKKKKKKKKKKR 62

RESULT 7
US-09-764-846-163
; Sequence 163, Application US/09764846
; Patent No. US20020102638A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT12
; CURRENT APPLICATION NUMBER: US/09/764,846
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 163
; LENGTH: 66
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE

```

```
; LOCATION: (4)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (56)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-846-163

Query Match          19.4%; Score 59; DB 10; Length 66;
Best Local Similarity 34.9%; Pred. No. 1;
Matches 15; Conservative 8; Mismatches 12; Indels 8; Gaps 1;

Qy 1 KVHGSARAG-----KVRGQTPKVAQOEKKKKKTGRKR 35
Db 20 RVRSSRSRAASMTFSKXXXXXXXXXXXXXXXXXXXXXK 62

RESULT 8
US-09-864-761-36199
; Sequence 36199, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wenheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 36199
; LENGTH: 72
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
```

```
; OTHER INFORMATION: MAP TO AC010826.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.7
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.8
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.2
US-09-864-761-36199

Query Match          19.1%; Score 58; DB 10; Length 72;
Best Local Similarity 41.5%; Pred. No. 1.5;
Matches 17; Conservative 6; Mismatches 12; Indels 6; Gaps 2;

Qy 4 GSLARAGKVRGQTPKVAQOEKKKKKT--GRAKRMOYNNR 41
Db 24 GKRKXKGR--RKVRKKEKKKTTRGRGRRRRRRR 61

RESULT 9
US-09-214-881A-2
; Sequence 2, Application US/09214881A
; Patent No. US20020009749A1
; GENERAL INFORMATION:
; APPLICANT: Ozaki, Shoichi
; APPLICANT: Sobajima, Junko
; APPLICANT: Uesugi, Hiroko
; APPLICANT: Okazaki, Takahiro
; APPLICANT: Tanaka, Masao
; APPLICANT: Nakao, Kazuwa
; APPLICANT: Yoshida, Michiteru
; APPLICANT: Shirakawa, Hitoshi
; APPLICANT: Osakada, Fumio
; TITLE OF INVENTION: DIAGNOSTIC DRUGS FOR AUTOIMMUNE DISEASES
; FILE REFERENCE: 068383.0104
; CURRENT APPLICATION NUMBER: US/09/214,881A
; CURRENT FILING DATE: 1999-06-07
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 208
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-881A-2

Query Match          19.1%; Score 58; DB 10; Length 208;
Best Local Similarity 34.9%; Pred. No. 4.6;
Matches 15; Conservative 7; Mismatches 17; Indels 4; Gaps 1;

Qy 19 VAKOEKKKKKTGRAKRMOYNNRFFNVVPTFGKKKG----PNA 57
Db 51 MSAKEKSKFEDMAKSDKARYDREMKNYVPPKGDKKKKKDPNA 93

RESULT 10
US-09-214-881A-6
; Sequence 6, Application US/09214881A
; Patent No. US20020009749A1
; GENERAL INFORMATION:
; APPLICANT: Ozaki, Shoichi
; APPLICANT: Sobajima, Junko
; APPLICANT: Uesugi, Hiroko
; APPLICANT: Okazaki, Takahiro
; APPLICANT: Tanaka, Masao
; APPLICANT: Nakao, Kazuwa
; APPLICANT: Yoshida, Michiteru
; APPLICANT: Shirakawa, Hitoshi
; APPLICANT: Osakada, Fumio
; TITLE OF INVENTION: DIAGNOSTIC DRUGS FOR AUTOIMMUNE DISEASES
; FILE REFERENCE: 068383.0104
```





